

UNIVERSIDADE FEDERAL DO PARANÁ

KAUÊ DE MORAES VESTENA

DESENVOLVIMENTO DE UM SISTEMA DE MAPEAMENTO MÓVEL TERRESTRE
DE BAIXO CUSTO PARA DETECÇÃO DA VEGETAÇÃO URBANA USANDO
REDES NEURAIS CONVOLUCIONAIS

CURITIBA

2020

KAUÊ DE MORAES VESTENA

DESENVOLVIMENTO DE UM SISTEMA DE MAPEAMENTO MÓVEL TERRESTRE
DE BAIXO CUSTO PARA DETECÇÃO DA VEGETAÇÃO URBANA USANDO
REDES NEURAIS CONVOLUCIONAIS

Dissertação de Mestrado apresentada como requisito parcial à obtenção do título de Mestre em Ciências Geodésicas, pelo Programa de Pós-Graduação em Ciências Geodésicas, do Setor de Ciências da Terra, da Universidade Federal do Paraná.

Orientador: Prof. Dr. Daniel Rodrigues dos Santos

CURITIBA

2020

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

V583d

Vestena, Kauê de Moraes

Desenvolvimento de um sistema de mapeamento móvel terrestre de baixo custo para detecção da vegetação urbana usando redes neurais convolucionais [recurso eletrônico] / Kauê de Moraes Vestena. – Curitiba, 2020.

Dissertação - Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas, 2020.

Orientador: Daniel Rodrigues dos Santos .

1. Vegetação – Mapeamento. 2. Vegetação urbana. 3. Redes neurais (Computação). 4. Sistemas de navegação. 5. Coordenadas geográficas. 6. Mapeamento móvel terrestre. I. Universidade Federal do Paraná. II. Santos, Daniel Rodrigues dos. III. Título.

CDD: 581.9

Bibliotecário: Elias Barbosa da Silva CRB-9/1894



MINISTERIO DA EDUCAÇÃO
SETOR DE CIÊNCIAS DA TERRA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS
GEODÉSICAS - 40001016002P6

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **KAUÊ DE MORAES VESTENA** intitulada: **DESENVOLVIMENTO DE UM SISTEMA DE MAPEAMENTO MÓVEL TERRESTRE DE BAIXO CUSTO PARA DETECÇÃO DA VEGETAÇÃO URBANA USANDO REDES NEURAIS CONVOLUCIONAIS**, sob orientação do Prof. Dr. DANIEL RODRIGUES DOS SANTOS, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

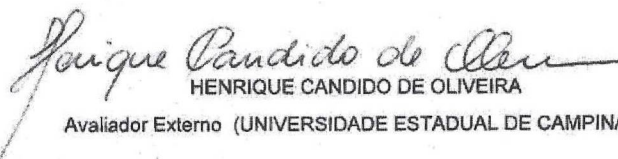
A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 31 de Janeiro de 2020.



DANIEL RODRIGUES DOS SANTOS

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)



HENRIQUE CANDIDO DE OLIVEIRA

Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)



HIDEO ARAKI

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Dedico este trabalho, com carinho, ao meu Pai.

Um dia, Ele me contou, usando um sotaque jocosamente afrancesado, que “um tal de *Lavoisier*” afirmou que “*na Natureza, nada se perde, nada se cria, tudo se transforma*”. De fato, Meu Pai se *transformou* na mais fortuita das lembranças, a qual jamais se *perderá*. Mas, Eu ousou discordar do cientista francês! Este mesmo, que de acordo com o meu Pai, “se ‘revirava’ no túmulo” a cada reação química não balanceada em provas de universitários... Discordo! Já que o legado de seu póstumo apreciador, indelevelmente, *criou* os mais primorosos frutos! Pobre *Antoine*... Toda vez que você lê... Enfim, Eu, ainda que sem crenças, prefiro dessa vez discordar do meu pai, e imaginar uma realidade onde os dois, e todos os demais estão, bebendo e conversando alegremente! E que cerveja um cientista consegue imaginar, senão aquela produzida por processos químicos, respeitando a Lei da Conservação das Massas?



In Memoriam,
Ao meu Cientista Favorito!

AGRADECIMENTOS

Aos meus pais, professora Mara e professor Mauro por tudo. Sobre tudo ao meu Pai, meu maior exemplo de Educador e Cientista, que sempre me incentivou ao cultivo do conhecimento em todas as suas formas, ao exercício do pensamento crítico, a ter consciência social...

À minha irmã Tainara e a todos os irmãos não-consanguíneos. A todos os meus amigos, inclusive os não humanos.

Ao professor Daniel, pela orientação e amizade, sempre compreensivo, permitindo liberdade no tema e no desenvolvimento da pesquisa.

À CAPES e ao CNPQ pelo fomento à pesquisa.

Às instituições UFPR e UTFPR pela formação acadêmica e introdução ao mundo das ciências geodésicas.

A toda a comunidade científica, a todos que desenvolvem e buscam conhecimento, sobre tudo ao conhecimento disponibilizado de maneira livre.

Às comunidades do *Software* e *Hardware* livres. Sem eles, este trabalho seria totalmente inviável. O “mundo de fora” tem muito a aprender com a forma de pensar e agir de tais comunidades.

Aos amigos do Laboratório de Aquisição e Processamento de Dados Espaciais, pelas horas de companhia e intercâmbio de conhecimento. Especialmente ao Alex B. Kraemer, que forneceu grande ajuda na montagem da plataforma física do SMMT apresentado neste trabalho.

Aos amigos do PPGCG, pelo suporte e companhia. Sobre tudo ao Marciano, pelo Fusca, cuja simbiose com a temática de baixo custo, é ímpar!

A você, leitor! Afinal, a sua leitura é a razão da existência do texto!

*“We must not forget that when radium was discovered no one knew that it would prove useful in hospitals. The work was one of pure science. And this is a proof that scientific work **must not** be considered from the point of view of the direct usefulness of it. It must be done for itself, for the beauty of science, **and then** there is always the chance that a scientific discovery may become like the radium a benefit for humanity.”*

Marie Skłodowska Curie, 1921

RESUMO

O processo de urbanização trouxe uma série de questões problemáticas como a poluição, microclimas abafados e progressiva impermeabilização. Tais problemas, que afetam diretamente a qualidade de vida da população das cidades, têm dentre os fatores importantes em sua mitigação a presença de vegetação nos cenários urbanizados. De tal modo, faz-se necessário o mapeamento da abrangência e do estado de saúde da assim intitulada "vegetação urbana", a fim de auxiliar o planejamento e a execução de políticas públicas que visem a extensão, o cuidado e também o manejo adequado, dado que a vegetação também pode causar danos quando em condições inadequadas. Uma tecnologia apresentada nas últimas décadas, capaz de fornecer grandes volumes de dados empregando pouco tempo de coleta é constituída pelos chamados Sistemas de Mapeamento Móvel Terrestre - SMMT, compostos primeiramente por um Sistema de Navegação e um Sistema de Imageamento, além de unidades de alimentação, controle síncrono e armazenamento. Ressalta-se que todos os componentes do SMMT estão embarcados em uma plataforma rígida, acoplável ao topo de um veículo terrestre. Nesse contexto, é proposta no presente trabalho a criação de um SMMT de baixo custo e que possui a importante característica de operação de imageamento na faixa espectral do Infravermelho próximo - NIR, cuja finalidade se deve à demanda por dados relativos à vegetação urbana. Em virtude da possibilidade de imageamento na faixa espectral do NIR, a vegetação é uma das feições com significativa discriminação nas imagens adquiridas pelo sistema em questão. Após a aquisição das imagens, torna-se necessária a separação temática da vegetação em relação à não-vegetação, obtendo-se como resultado os pixels associados a cada uma dessas duas classes, num processo conhecido como segmentação semântica. A gama de algoritmos que se tornou o estado-da-arte no tocante à segmentação semântica ("classificação") é composta pelas Redes Neurais Convolutivas - CNNs, que permitem resolver tal questão. As CNNs são treinadas por algoritmos de Deep Learning, tendo como entradas para o treinamento amostras com resultado conhecido e amostras ocultas para validação e aprendizado. No presente trabalho, foram treinadas e testadas redes com 50 imagens nas quais a vegetação foi segmentada manualmente, atingindo-se uma média de 83% para o indicador *Intersection Over Union* - IoU (razão entre Interseção e União), ainda que a rede não desempenhe da maneira desejada na ausência de vegetação, fator mitigado com a utilização de uma segunda rede, sendo esta do tipo multi-classe. A partir dos resultados obtidos, que permitiram atestar bons desempenhos para a rede neural treinada, é possível concluir que o SMMT criado (que é totalmente operacional) está apto para capturar dados sobre a vegetação urbana.

Palavras-chave: Sistemas de Mapeamento Móvel Terrestre; Imageamento Multi-Espectral; Vegetação Urbana; Redes Neurais Convolutivas;

ABSTRACT

The urbanization process brought a lot of problematic issues, like pollution, unpleasant microclimates and progressive surface imperviousness. These problems, which directly affect the urban population's life quality, can be mitigated by the presence of urban vegetation. Therefore, it is necessary to map their coverage and health status globally in urban areas, in order to assist the planning and implementation of public policies aimed at extension, care and also adequate management, since vegetation can also cause damage in inappropriate conditions. As a technology presented in the last decades, the so-called Terrestrial Mobile Mapping Systems - TMMS, are capable of providing large amounts of data without lots of time consumption, they are composed primarily by a Navigation System and an Imaging System, in conjunction with Power Supply, Synchronous Control, and Storage Units. It is important to note that all TMMS components are mounted on a rigid platform, attachable to the top of a ground vehicle. In this context, is proposed the creation of a low-cost TMMS, which has the distinctive feature of imaging in the near-infrared (NIR) spectrum, whose purpose is due to the demand for urban vegetation data. Because of the possibility of imaging in the NIR band, the vegetation is among the features with significant discrimination in the images acquired by the proposed system. After the image acquisition step, it becomes necessary the thematic separation of vegetation from non-vegetation, obtaining, as a result, the pixels associated with each of these two classes, this process is known as semantic segmentation. The set of algorithms that became the state of the art, in semantic segmentation scope, are composed essentially by Convolutional Neural Networks - CNNs. This kind of network is trained by Deep Learning algorithms, using as inputs a set of samples with known results (for training) and hidden samples for evaluation and improvement. In this study, CNNs were trained and tested with 50 images, in which vegetation was manually segmented, reaching a mean value of 83% for the Intersection Over Union (IoU) indicator, although the net does not perform well in the absence of vegetation, a factor mitigated with the use of a second net, that was of the multi-class type. From the results obtained, which demonstrated good performance for the trained neural network, it is possible to conclude that the developed TMMS (which is fully operational) is adequate to capture data regarding urban vegetation.

Keywords: Terrestrial Mobile Mapping Systems; Multi-Spectral Imaging; Urban Vegetation; Convolutional Neural Networks;

LISTA DE FIGURAS

Figura 1.1: cena urbana com alguns elementos identificados	25
Figura 1.2: imagem de satélite de plantação e NDVI calculado a partir desta	26
Figura 1.3: O Sistema MMS-G220, embarcado em plataforma veicular.	28
Figura 1.4: Sensores utilizados em um carro autônomo e seu custo	29
Figura 1.5: Evolução do Estado-da-Arte em Segmentação Semântica	30
Figura 2.1: Sistemas Geodésicos de Coordenadas	33
Figura 2.2: BF de uma câmera e referencial. digital	34
Figura 2.3: Geometria do Posicionamento por Satélites	36
Figura 2.4: Fluxograma do Funcionamento de um INS	37
Figura 2.5: O Espectro Eletromagnético	39
Figura 2.6: Fluxograma de transformação de Luz em Imagens Digitais	39
Figura 2.7: O Filtro de Bayer	40
Figura 2.8: Curvas de Sensibilidade Espectral envolvendo sensores e filtros	41
Figura 2.9: Exemplo de respostas espectrais e valores de NDVI para folhagens	43
Figura 2.10: NDVI Calculado para toda a superfície terrestre	43
Figura 2.11: GEOVAN, o SMMT do ICC	45
Figura 2.12: O SMMT Comercial da Mitsubishi	46
Figura 2.13: Esquematização da estrutura de uma CNN simples	48
Figura 2.14: Ilustração do processo de Retropropagação	49
Figura 2.15: Categorias de aplicações genéricas de CNNs em classificação	50
Figura 2.16: Exemplo da Estrutura Básica de uma Rede Completamente Convolutacional	50
Figura 3.1 O Protótipo do SMMT Proposto e seus elementos	53
Figura 3.2: Fluxograma das etapas do método	56
Figura 3.3: Esquema de controle e gravação de dados do sistema apresentado	60
Figura 3.4: O SMMT Instalado em Plataforma Veicular Terrestre	61
Figura 3.5: Interface utilizada para acionamento dos sensores da plataforma	62
Figura 3.6a : Trajetória percorrida pelo veículo em um teste de exemplo	63
'Figura 3.6b: gráficos com dados extraídos da IMU	64
Figura 3.7: Amostra de imagens tomadas pela plataforma	65

Figura 3.8: esquematização do uso de CNNs no presente trabalho	66
Figura 3.9: Pares de Imagens disponibilizados no Cityscapes Dataset	68
Figura 3.10: Amostras de máscaras de classificação de imagens tomadas pela plataforma	69
Figura 3.11: Possibilidades para a comparação de um pixel predito e da amostra de controle	71
Figura 3.12: Exemplo ilustrativo de predição sobreposta a imagem de controle	72
Figura 3.13: demonstração conceitual da métrica “IoU”	74
Figura 3.14: Exemplo da evolução da Taxa de Acerto (TA)	75
Figura 3.15: Fotografia e imagens geradas a partir dela e da classificação	78
Figura 4.1: Detalhes da trajetória GNSS sobreposta a ortoimagem do Bing	83
Figura 4.2: Gráfico da Evolução da Métrica IoU para o Cenário A-I	85
Figura 4.3: Gráfico da Evolução da Métrica IoU para o Cenário A-II	86
Figura 4.4: Gráfico da Evolução da Métrica IoU para o Cenário B	86
Figura 4.5: Box-Plots das métricas para a melhor época do cen. A-I	88
Figura 4.6: Box-Plots das métricas para a melhor época do cen. A-II	89
Figura 4.7: Box-Plots das métricas para a melhor época do cen. B	89
Figura 4.8: Métricas de erro individuais da CNN escolhida	92
Figura 4.9: Amostras de Resultados da Classificação Realizada	93
Figura 4.10: Demonstração do resultado gerado para a CNN Final em imagens sem vegetação	95
Figura 4.11: Demonstração do resultado gerado para a CNN Final em imagens sem vegetação	96
Figura 4.12: Mudança nos tons de cinza nas mesmas feições, imagens diferentes	98
Figura 4.13: histograma e imagem com valores típicos de NDVI calculados	98
Figura 4.14: variação dos valores calculados de NDVI	99
Figura A1.1: exemplos de aproximações monoscópicas de ortofotos geradas com as imagens	111
Figura A1.2: detalhes de uma trajetória gerada com as imagens do SMMT	112
Figura A2.1: uma típica imagem tomada com o SMMT	114
Figura A2.2: Detalhe da realização das medidas de referência e do diâmetro na imagem	115

LISTA DE TABELAS

Tabela 1.1: critérios para plantio de árvores em São Paulo - SP	22
Tabela 1.2: Evolução do Estado-da-Arte em Segmentação Semântica	31
Tabela 3.1 Custos individuais dos equipamentos do protótipo do SMMT	54
Tabela 4.1: Parâmetros de captura dos dados do SMMT utilizados no experimento	80
Tabela 4.2: resumo dos cenários analisados para redes neurais	81
Tabela 4.3: valores de estatística descritiva da validação das melhores épocas	90
Tabela A2.1: Medidas indiretamente obtidas por meio de relações de escala	116

LISTA DE SIGLAS E ACRÔNIMOS

BF	<i>Body Frame</i> (Sistema de Coordenadas Próprio de um “Corpo ¹ ”)
CCD	<i>Charged Coupled Device</i> (Dispositivo de Carga Acoplada)
CMOS	Complementary Metal-Oxide-Semiconductor, (Semicondutor de Metal-Óxido Complementar)
CNN	<i>Convolutional Neural Network</i> (Rede Neural Convolutiva);
CSV	<i>Comma-Separated Values</i> (Valores Separados por Vírgula);
DOP	<i>Dilution Of Precision</i> (Diluição da Precisão);
ECEF	<i>Earth-Centered, Earth-Fixed</i> (referencial Centrado na Terra e Afixado à Terra ²)
FN	Falso Negativo;
FP	Falso Positivo;
GNSS	<i>Global Navigation Satellite System</i> (Sistema Global de Navegação por Satélites)
INS	<i>Inertial Navigation System</i> (Sistema de Navegação Inercial)
IoU	<i>Intersection over Union</i> (Razão da Interseção pela União);
MVC	Matriz Variância-Covariância;
NDVI	<i>Normalized Difference Vegetation Index</i> (Índice de Vegetação da Diferença Normalizada);
NIR	<i>Near Infrared</i> (Infravermelho Próximo);
RBMC	Rede Brasileira de Monitoramento Contínuo;
R/NIR	<i>Red / NIR</i> (Vermelho / NIR);
SMMT	Sistema de Mapeamento Móvel Terrestre;
SGL	Sistema Geodésico Local
TA	Taxa de Acerto;
VP	Verdadeiro Positivo;
VN	Verdadeiro Negativo;

¹ Um corpo Rígido e no contexto do presente trabalho é geralmente um sensor.

² Tradução Livre.

SUMÁRIO

1	INTRODUÇÃO	17
1.1	CONSIDERAÇÕES GERAIS	17
1.1	OBJETIVO GERAL	19
1.2	OBJETIVOS ESPECÍFICOS	19
1.3	MOTIVAÇÃO	20
1.4	RELEVÂNCIA PARA A CIÊNCIA E ECONOMIA	26
1.6	ESTADO-DA-ARTE	27
1.6.2	Estado da Arte em Sistemas de Mapeamento Móvel Terrestre	28
1.6.2	Estado da Arte em CNNs para Segmentação Semântica	29
1.5	PREÂMBULO E ESTRUTURA DO TRABALHO	31
2	FUNDAMENTAÇÃO TEÓRICA	33
2.1	SISTEMAS DE NAVEGAÇÃO E SISTEMAS DE COORDENADAS	33
2.1.2	Quatérnios Unitários Como Parametrização da Orientação 3D	34
2.2	SISTEMAS DE POSICIONAMENTO GLOBAL POR SATÉLITES	35
2.3	SISTEMAS DE NAVEGAÇÃO INERCIAL	37
2.4	O ESPECTRO ELETROMAGNÉTICO E SENSORES PASSIVOS	38
2.4.2	Imageamento R-NIR com Aplicação de Filtro Vermelho	39
2.4.2	NDVI	42
2.5	SISTEMAS DE MAPEAMENTO MÓVEL TERRESTRE	44
2.6	REDES NEURAIS CONVOLUCIONAIS E DEEP LEARNING	46
3	MATERIAIS E MÉTODOS	51
3.1	MATERIAIS E RECURSOS	51
3.2	MÉTODO	55
3.2.2	Visão Geral Do Método	55
3.2.2	Aquisição dos dados	58
3.2.2.1	Captura de Dados	58
3.2.2.2	Coleta de Dados	60
3.2.2	Processamento dos dados	64
3.2.3.1	Treinamento e Uso das Redes Neurais Convolucionais	65
3.2.3.2	Processamento dos Dados Obtidos pelo SMMT	75
4	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	80

4.1	EXPERIMENTOS REALIZADOS	80
4.2	RESULTADOS E ANÁLISES	81
4.2.2	ANÁLISE DOS DADOS DE NAVEGAÇÃO	82
4.3	ANÁLISE DOS RESULTADOS DE SEGMENTAÇÃO PELA CNN	84
4.3.2	Análise da classificação em Fotografias sem Vegetação	94
4.4	ANÁLISE DA VALIDADE DOS VALORES DE NDVI CALCULADOS	97
5	CONSIDERAÇÕES FINAIS	101
5.1	SUGESTÕES PARA TRABALHOS FUTUROS	102
	REFERÊNCIAS	105
	APÊNDICE 1 - OUTRAS POTENCIALIDADES DO SISTEMA PROPOSTO	110
	APÊNDICE 2 - OBTENÇÃO SIMPLES DE MEDIDAS DE ÁRVORES	113
	APÊNDICE 3 - ETAPAS NÃO IMPLEMENTADAS OU NÃO REPORTADAS	117
	APÊNDICE 4 - SOBRE O TEMA ANTIGO E SEUS PROBLEMAS TÉCNICOS	121
	APÊNDICE 5 - MAPA PRODUZIDO COM OS DADOS PROCESSADOS	123

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES GERAIS

A urbanização é um processo que se confunde com a história da própria civilização³. Trata-se de um processo que, devido a diversos fatores, produz cenários com uma progressiva aglomeração de indivíduos. Isto torna a vida destes marcada por uma série de questões relacionadas a problemas específicos do meio urbano, como a exposição à poluição e à vulnerabilidade a enchentes.

Nesse âmbito, a vegetação urbana exerce importante papel na manutenção da qualidade de vida nesses diversos ambientes. Uma vez que ela, possui a capacidade de mitigar alguns dos problemas gerados pela urbanização. Assim sendo, faz-se mister o contínuo ato de responder à seguinte questão: considerando-se toda uma cidade, quais ruas estão adequadamente arborizadas?

A atividade de mapeamento possui a real capacidade de fornecer a resposta a tal pergunta. Trata-se de uma atividade fundamental para qualquer área de planejamento estratégico, ainda que, para diversos temas seja comum a falta ou mesmo inexistência de mapeamento, como normalmente é o caso da vegetação urbana.

A paisagem urbana geralmente se modifica com acentuada frequência, tal característica é ainda maior na vegetação. Assim, apresenta-se a demanda por uma metodologia para coleta de tais dados espaciais com algumas características específicas: viabilidade econômica; capacidade para realização de coletas com alta frequência; e opcionalmente, a independência de recursos externos.

De acordo com EL-SHEIMY (2005) os chamados *Mobile Mapping Systems*, ou SMMT - Sistemas de Mapeamento Móvel Terrestre⁴, atendem a tais requisitos⁵, sobretudo devido à alta frequência, pois são montados em plataformas veiculares terrestres que podem cobrir grandes áreas em pouco tempo, além de dificilmente

³ Fonte: **Uruk: a primeira cidade.** Disponível em: <<https://super.abril.com.br/historia/uruk-a-primeira-cidade/>>.

⁴ Tradução que aparece nacionalmente, como em VESTENA, 2017.

⁵ Ainda que possa se abrir mão de alguns destes requisitos para se obter características específicas: maior custo pode elevar a qualidade dos dados de saída; dependência de bases GNSS, possibilita posicionamento relativo, etc.

violar qualquer norma de segurança existente. É possível elaborar sistemas de baixo custo, viabilizando a atuação de múltiplos agentes⁶, contribuindo assim, ainda mais para uma alta frequência de atualização. A independência de recursos externos favorece uma opção pelo baixo custo, ainda que, em algumas áreas, seja possível aproveitar a infraestrutura pública⁷.

Ainda de acordo com o autor, um SMMT é composto por: uma plataforma rígida; um sistema de controle e alimentação; um sistema de imageamento; e um sistema de navegação. Há grande liberdade na escolha específica dos componentes de tal amálgama de sistemas, podendo se optar por diversas faixas de preço, geralmente havendo uma relação direta entre o custo⁸ e a qualidade dos dados fornecidos, entretanto, isso pode ser compensado por outras possíveis vantagens.

Assim sendo, reitera-se a afirmação que um SMMT de baixo custo possui capacidade para atender a demanda por mapeamento da vegetação urbana, ao servir de sistema de coleta para os dados brutos usados para geração de informação geoespacial ligada a tal tema. Tais dados, em uma categorização mais ampla, podem ser divididos em duas categorias: navegação e imageamento. A primeira costuma usufruir de metodologias bem estabelecidas para se chegar ao resultado desejado.

Com relação ao imageamento, há uma gama muito mais ampla de possíveis cadeias de processamento a serem adotadas (em junção ou não com a navegação) para se chegar a resultados também diversificados. Entretanto, quando objetiva-se um tema específico, e as imagens capturam diversos objetos, torna-se imprescindível separar o tema de interesse dos demais, já que estes últimos (carros, pedestres, pavimentação...) não serão objeto de análise posterior. Nesse âmbito, a tarefa de classificar vegetação e não-vegetação será menos custosa caso seja realizada diretamente nas fotografias. A classificação em imagens digitais, parte da tarefa de se atribuir uma classe para cada pixel (LONG et. al., 2015).

⁶ Termo Genérico, referindo-se a qualquer pessoa ou máquina que opere um sistema capaz de mapear.

⁷ Como no caso das cidades com estações da RBMC, quando tecnicamente viável.

⁸ Porém, com o advento de tecnologias criadas com *hardware* livre, o custo de diversos equipamentos tem reduzido.

Há diversas técnicas disponíveis para a realização das tarefas de segmentação/classificação. De acordo com vários autores, como CHEN et al, (2014), MOSTAJABI et. al. (2015), BADRINARAYANAN et. al. (2017), os melhores resultados em tal área (estado-da-arte) estão sendo obtidos por meio das denominadas Redes Neurais Convolucionais (CNNs). Tais redes são treinadas por meio de algoritmos de *Deep Learning*, que geram um modelo (uma rede) capaz de modelar classificações com regras adaptáveis, baseados em uma série de amostras com resultados conhecidos⁹, isto é, amostras de referência.

Assim sendo, decidiu-se pelo treinamento de CNNs para a realização da já mencionada tarefa indispensável de classificar vegetação e não vegetação, de modo que os dados do SMMT a partir deste ponto estejam aptos a serem utilizados para a geração de informação geoespacial passível de ser utilizada para análises da vegetação urbana.

1.1 OBJETIVO GERAL

Criar e desenvolver um Sistema de Mapeamento Móvel Terrestre de baixo custo cujas imagens possuam características adequadas ao estudo de vegetação urbana.

1.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho são:

- I. Escolher os componentes para montagem do SMMT;
- II. Estabelecer a metodologia de coleta e processamento dos dados capturados pelo SMMT em um levantamento;
- III. Treinar uma Rede Neural Convolutiva (CNN) capaz realizar a segmentação semântica (classificação) de vegetação e não vegetação nas imagens tomadas pelo imageador do SMMT; e
- IV. Avaliar a acurácia da classificação com utilizando amostras de controle (classificadas manualmente).

⁹ Geralmente tais amostras são classificadas manualmente.

1.3 MOTIVAÇÃO

Ao se idealizar a presente pesquisa, pensou-se em um tema que envolvesse o interesse público. A vegetação urbana, cuja importância já foi comentada em 1.1, é fundamental para a vida nas cidades, conhecê-la adequadamente é relevante para a elaboração de políticas visando sua manutenção, expansão e manejo adequado. Muitas administrações municipais não possuem recursos para a realização do monitoramento da condição de sua vegetação, faltando assim dados para seu contínuo monitoramento e fiscalização.

De acordo com NICODEMO & PRIMAVESI (2009), as plantas têm grande papel no bem estar humano, pois desencadeiam reações psicológicas positivas, trazem sensação de conforto emocional e abrigo, favorecem a saúde mental, e, em alguns casos, houveram evidências de que proporcionam melhorias na saúde física. Os autores apresentam estudos onde a presença de árvores aponta ainda correlação com a redução da violência, sobretudo a doméstica.

De acordo com os autores supra referidos, há seis grandes tópicos relacionados à relevância da vegetação em meios urbanos:

1. Microclima: atua na amenização da temperatura (sobretudo de calor, mas também para o frio) e umidificação do ar, o que reduz a ocorrência de ilhas de calor e propicia sombras que também sombreiam o asfalto, de modo a reduzir a absorção de calor e reverberação pelo mesmo;
2. Redução da poluição do ar: os estômatos¹⁰ absorvem para a superfície das folhas uma série de gases poluentes. A folhagem também é responsável por reter partículas sólidas, devido à convecção que leva o ar mais carregado para baixo. Estima-se que em 1991, na cidade de Chicago, a arborização urbana removeu 5575 t de poluentes, gerando uma economia de US\$ 9.2 milhões para o poder público, o estudo evidenciou também que a remoção de poluentes cresce exponencialmente com o tamanho da árvore;

¹⁰ São estruturas presentes na superfície das folhas, responsáveis pela comunicação do meio interno com a atmosfera (Fonte: www.sciencedirect.com).

3. Redução do ruído: as plantas atenuam a intensidade do som pela reflexão (espalhamento) da energia das ondas sonoras e pela absorção da energia das mesmas. A folhagem também produz um agradável som de fundo (*white noise*) devido ao seu farfalhar, que reduz a percepção de outros sons. É possível intercalar vegetação alta e rasteira para produzir barreiras mais efetivas contra o som, sendo que mesmo a grama é mais eficiente que outras pavimentações em tal aspecto;
4. Interceptação da água da chuva: o crescimento das raízes e a deposição de matéria orgânica pelas árvores aumentam a taxa e capacidade de infiltração da água no solo. As copas das árvores também protegem o solo do impacto das gotas de chuva, conservando melhor a permeabilidade do mesmo, além de retardarem o escoamento superficial, de modo a retardar o pico de uma enchente, considerando ainda que a própria árvore vai absorver parte da água. A transpiração da vegetação lança água do solo para a atmosfera, o que aumenta ainda mais o potencial de armazenamento de água no solo, tal processo também melhora a condição da umidade do ar, que beneficia a população, sobretudo a suscetível a problemas respiratórios. Estima-se que uma árvore de médio porte é capaz de interceptar 12795 L de água por ano;
5. Retenção de carbono: as árvores urbanas, assim como as demais, atuam na redução da disponibilidade atmosférica de gases do efeito estufa devido a seu processo respiratório, superavitário na absorção de gás carbônico. Estima-se que cada árvore grande sequestre cerca de 93 Kg de dióxido de carbono por ano;
6. Valores estéticos e culturais: em alguns contextos urbanos, as árvores agregam valor monetário aos imóveis das redondezas, sendo que algumas prefeituras dão incentivos para a presença de árvores e solo permeável nos lotes urbanos. O estabelecimento de jardins e hortas comunitárias favorece a interação social e a construção de um senso de comunidade;

Apesar de todos os benefícios oferecidos pela vegetação urbana, há uma série de danos em potencial causados por eventuais quedas. Geralmente são danos patrimoniais, envolvendo a rede elétrica, muros e carros. Com a ocorrência de uma sequência grande de ventos associados a chuva, o solo se torna instável devido ao seu encharcamento. Esse risco é potencializado quando o plantio foi realizado sem planejamento, envolvendo espécies exóticas que não se adaptam bem ao solo local, assim como devido ao dimensionamento inadequado de canteiros, calçadas e ações de manejo mal planejadas, como podas que favorecem o apodrecimento de troncos e raízes. Uma vez que as árvores são seres vivos, se faz necessário acompanhar seu estado de saúde (adaptado de BRAZOLIN, 2015).

Apresentados essas duas grandes facetas da arborização urbana, percebe-se que precisam ser conduzidas políticas públicas para planos de plantio e manejo, que, idealmente, devem seguir uma série de parâmetros, como os requeridos pelo município de São Paulo, apresentados na tabela 1.1.

Tabela 1.1: critérios para plantio de árvores em São Paulo - SP

Distância mínima em relação à	Características Máximas da Espécie		
	Pequeno porte	Médio porte	Grande porte
Esquina (referenciada ao ponto de encontro dos alinhamentos dos lotes da quadra em que se situa)	5,0m	5,0m	5,0m
Iluminação pública	(1) ¹¹	(1)	(1) e (2)
Postes	3,0m	4,0m	5,0m e (2)

¹¹(1) : Evitar interferências com cone de iluminação; (2): Sempre que necessário, a copa de árvores de grande porte deverá ser conduzida acima das fiações elétricas e da iluminação pública; (3): A visão dos usuários não deve ser obstruída; (4) Caso as espécies arbóreas sejam diferentes pode ser adotada a média aritmética; (5) Uma vez e meia o raio da circunferência circunscrita à base do tronco da árvore quando adulta, medida em metros;

Distância mínima em relação à	Características Máximas da Espécie		
	Pequeno porte	Médio porte	Grande porte
Equipamentos de segurança (hidrantes)	1,0m	2,0m	3,0m
Instalações subterrâneas (gás, água, energia, etc.)	1,0m	1,0m	1,0m
Ramais de ligações subterrâneas	1,0m	3,0m	3,0m
Mobiliário urbano (bancas, cabines, guaritas, telefones)	2,0m	2,0m	3,0m
Galerias	1,0m	1,0m	1,0m
Caixas de inspeção (boca-de-lobo, boca-de-leão, poço-de-visita, bueiros, etc.)	2,0m	2,0m	3,0m
Fachadas de edificação	2,40m	2,40m	3,0m
Guia rebaixada, gárgula, borda de faixa de pedestre	1,0m	2,0m	1,5R (5)
Transformadores	5,0m	8,0m	12,0m
Espécies arbóreas	5,0m (4)	8,0m (4)	12,0m (4)

FONTE: Prefeitura Municipal de São Paulo - SP

Com relação à demanda por tais dados, obtê-los trata-se de um grande desafio. Sua coleta é um problema que se superdimensiona com o contínuo crescimento das cidades. Há técnicas tradicionais de coleta, executadas por meio de agentes que obtêm informações a respeito da vegetação, preenchendo formulários e realizando medições *in situ*. Tais técnicas possuem baixa produtividade diária, assim, utilizá-las para o estabelecimento de uma base completa e com atualização frequente, dependeria de grande dispêndio de recursos. Curitiba, por exemplo, possui em torno de 4515 Km¹² de malha viária urbana, algo desafiador.

¹²

Fonte: <https://bandnewsfmc Curitiba.com/quase-metade-da-malha-viaria-de-curitiba-passou-por-processo-de-tapa-buraco-nesse-ano/> > .

Assim, objetivando propor uma solução que viabilize tal tipo de mapeamento, buscou-se empregar uma tecnologia que possibilite uma coleta capaz de gerar uma base de dados com grande completude e agilidade inatingível por outros meios.

As características supracitadas só podem ser alcançada sem trazer novas preocupações com segurança¹³, por meio do uso de sistemas imageadores aliados a sistemas de navegação embarcados em plataformas veiculares terrestres (carros). E esses são os componentes centrais de SMMTs, capazes de atender à demanda da aquisição de dados sobre vegetação urbana.

Há SMMTs baseados em dois tipos de imageadores¹⁴: ativos ou passivos. Os imageadores da segunda categoria, são basicamente câmeras fotográficas. Tais câmeras destacam-se pelo baixo custo, possuindo características geométricas e radiométricas especiais.

As fotografias capturam em uma projeção bidimensional, em perspectiva, a geometria da cena tridimensional, de modo a ser possível recuperar tal geometria por meio de técnicas de matemática aplicada (HARTLEY, 2003). Fotografias reais são tomadas por câmeras reais, estas que possuem como elementos principais um sistema de lentes e um sensor fotográfico.

É importante destacar a riqueza de detalhes presente em uma imagem fotográfica, sobretudo em meios urbanos, a imagem vai apresentar elementos, sejam fixos ou móveis, dos mais diversificados: calçadas, pavimentação, construções, placas, postes, carros, pedestres, e, evidentemente, a vegetação. Na figura 1.1 está apresentada uma cena urbana com alguns elementos identificados.

¹³ Há plataformas consideradas disruptivas como os *drones*, que ainda desafiam questões de segurança para seu emprego na coleta contínua e eventualmente autônoma de dados.

¹⁴ Alguns possuem ambos.

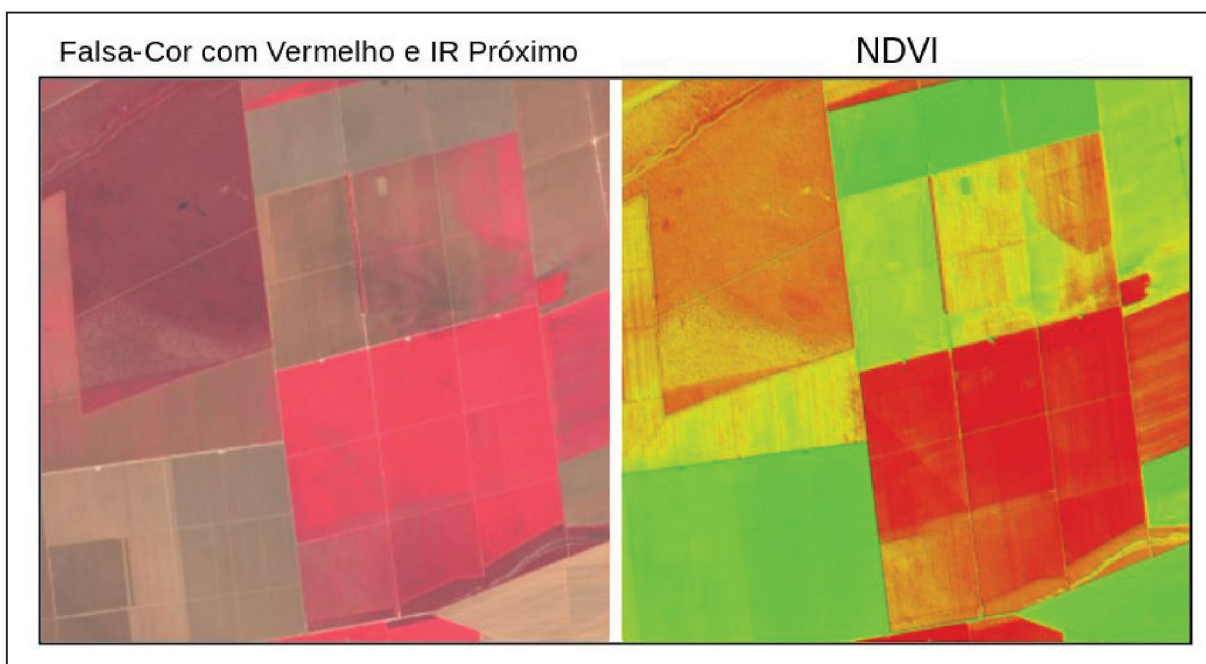
Figura 1.1: cena urbana com alguns elementos identificados



FONTE: <www.mappilary.com>

Tal sensor “foto-gráfico”, como trivialmente notável em sua denominação, também captura características radiométricas, isto é, da reflectância que os objetos imageados apresentam a faixas do espectro eletromagnético, as quais o objeto foi exposto e também estejam na faixa a qual o imageador é sensível. Tais reflectâncias ou *respostas espectrais*, estão diretamente ligadas a propriedades físico-químicas dos materiais que compõem os objetos imageados, assim, por meio delas, é possível inferir a respeito da natureza ou do estado dos objetos imageados (CARTER & KNAPP, 2001). Na figura 1.2 está apresentado um exemplo de uso de características radiométricas para se calcular um índice que está diretamente ligado ao estado de saúde dos indivíduos (“plantas”) numa área de cultivo.

Figura 1.2: imagem de satélite de plantação e NDVI calculado a partir desta

FONTE: <<http://www.engesat.com.br>>

Assim sendo, nota-se que a categoria de sistema proposto possui grande potencial para fornecer os dados que atendam as demandas aqui expostas. Tal potencial é ainda elevado pela proposição de que o sistema seja materializado como uma solução de baixo custo, com todo seu projeto existente, tanto a nível de *software*, quanto de especificações de *hardware*, de maneira totalmente livre. Assim, não haverá qualquer impeditivo para a continuidade de seu desenvolvimento e uso para qualquer interessado.

Voltando às questões relacionadas à riqueza de detalhes em fotografias, tal questão serviu de motivação para emprego de uma técnica de segmentação capaz de separar na imagem os pixels que são vegetação dos demais. Então, como já enunciado em 1.1, decidiu-se pelo uso das Redes Neurais Convolutivas para a realização de tal tarefa, que deverá preceder as demais, no tocante a mapeamento.

1.4 RELEVÂNCIA PARA A CIÊNCIA E ECONOMIA

Do ponto de vista econômico, o desenvolvimento de todo tipo de equipamento de baixo custo possui uma série de fatores positivos para sua aplicação em

situações reais, pois: possibilita que um número maior de unidades opere com o mesmo orçamento; reduz os danos de eventuais perdas, independentemente da causa; aumenta a independência com relação a soluções proprietárias tanto em *hardware* quanto em *software*, o que reduz a dependência de técnicos autorizados para manutenção.

Todas as ferramentas empregadas no presente trabalhos são de uso livre¹⁵, sendo de código aberto em sua maioria. O uso de padrões livres, tanto no campo de software quanto de hardware, possibilita uma comunicação eficiente entre diferentes componentes, uma vez que não há entraves advindos de “fórmulas secretas”, que frequentemente dificultam a integração de diferentes sistemas.

O mapeamento fornece informações para planejamento de tomada de decisões, permitindo ao poder público agir de maneira preventiva, estratégia que tende a ser muito mais barata que com medidas reparatórias. Como já exposto em 1.3, há grande relevância no monitoram

ento da vegetação urbana tanto em termos quantitativos quanto qualitativos.

Trata-se de um trabalho de aplicação do conhecimento científico. Assim sendo, do ponto de vista da ciência, o presente trabalho é uma amostra do conjunto de trabalhos do campo da engenharia. Tal área retorna para a sociedade contribuições na solução de problemas enfrentados pela mesma, contribuindo assim para fortalecer a importância da ciência diante da sociedade.

1.6 ESTADO-DA-ARTE

Esta seção dedica-se a fornecer ao leitor uma visão genérica do estado presente (2019) nos dois tópicos centrais para presente trabalho: SMMTs e CNNS. Para uma apresentação mais formal a respeito de tais tópicos, o leitor deverá verificar as seções 2.5 e 2.6, respectivamente.

¹⁵ Todas o são para uso acadêmico. A grande maioria é livre para qualquer uso.

1.6.1 Estado da Arte em Sistemas de Mapeamento Móvel Terrestre

Os SMMT já são uma tecnologia consagrada no mercado de equipamentos de mensuração, tendo como atrativos divulgados, características como “capacidade de realizar até um milhão de medidas por segundo”; sendo que várias marcas oferecem sistemas prontos para aquisição, dentre elas: *Trimble*; *Topcon*; *Leica*; *Teledyne*; e a chinesa *HI-TARGET* (CHAKRAVARTY, 2018).

Tais sistemas já são uma grande tendência para o futuro, dado que tal tipo de sistema é o “bloco construtivo” básico de um carro autônomo, com a importante distinção que o SMMT para tais carros precisa gerar resultados em tempo real e com latência mínima (TOTH, et. al., 2007).

Assim sendo, o estado-da-arte em desenvolvimento de tecnologias de “Mapeamento Móvel Terrestre” e “Carros autônomos” estão intrinsecamente ligadas, uma vez que a segurança de navegação autônoma depende de um mapeamento acurado, tanto de feições “permanentes” quanto de entidades móveis.

Um exemplo de sistema voltado para a aquisição de dados-base para sistemas operantes em tempo real é o *Mitsubishi MMS-G220*, projetado para coleta de dados para a olimpíada de Tóquio de 2020. O sistema a 40 Km/h é obtém uma acurácia absoluta na ordem de 10 cm para objetos a 7 m dos sensores de varredura *laser* embarcados (BOYD, 2016). O referido sistema está exibido na figura 1.3.

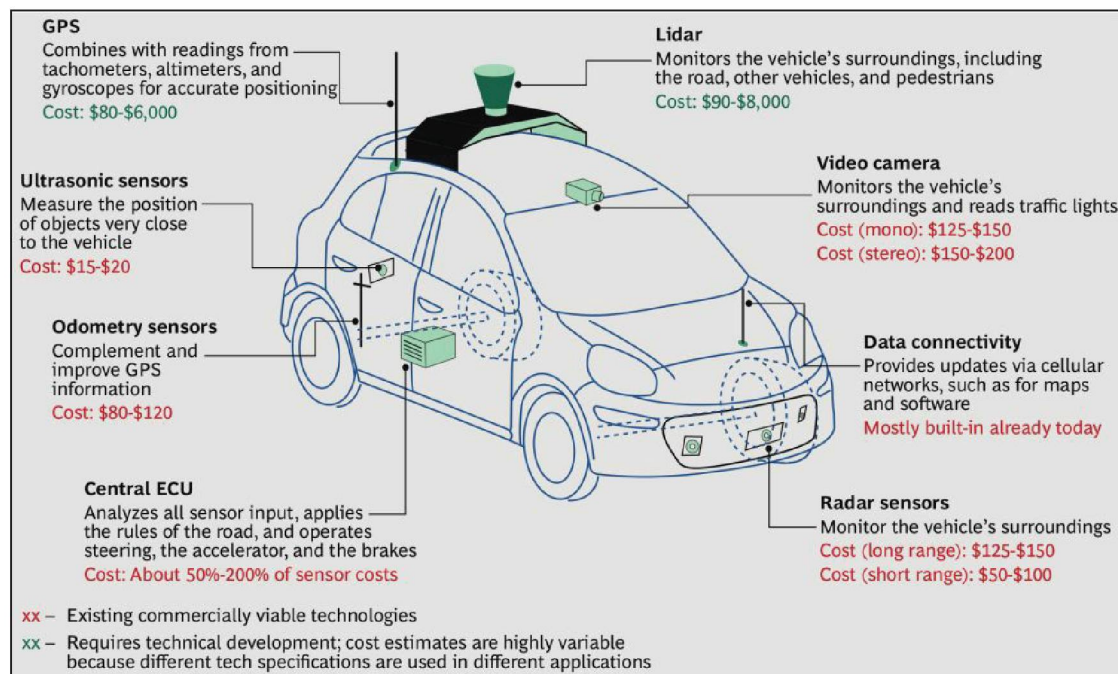
Figura 1.3: O Sistema MMS-G220, embarcado em plataforma veicular.



FONTE: BOYD (2016)

Na figura 1.4 está apresentada uma esquematização de um carro autônomo, apresentado por (ALHASHIMI, 2018), assim como os custos estimados para a cidade de Boston, nos EUA.

Figura 1.4: Sensores utilizados em um carro autônomo e seu custo



FONTE: (ALHASHIMI, 2018) e Boston Consulting Group

Como o leitor pode notar, a distinção entre SMMTs e algo como uma “rede de sensores” para carros autônomos é cada vez menos observável. Outra questão, perceptível na figura 1.4, que vêm de encontro com uma temática central ao presente trabalho, é a questão do custo: tem-se uma progressiva preocupação com sua redução, visando tornar as tecnologias empregadas cada vez mais acessíveis.

1.6.2 Estado da Arte em CNNs para Segmentação Semântica

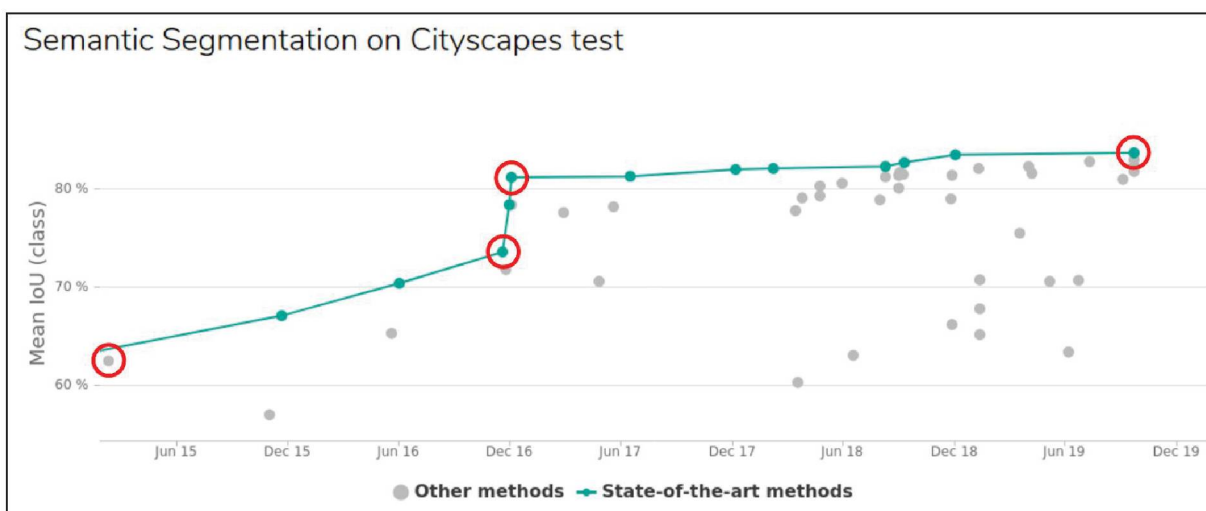
Ao contrário de muitos temas abordados ao longo do presente trabalho, a avaliação da qualidade de uma “Classificação de Imagens”, ou “Segmentação Semântica” pode ser estabelecida de maneira expressamente objetiva: por meio de um número percentual considerado como sendo a métrica de erro, diretamente

ligada ao nível de acurácia obtida pelo algoritmo utilizado. Geralmente é considerado o valor médio ao longo das classes consideradas, como será apresentado.

Dada tal possibilidade de se estabelecer uma qualificação quantificável, torna-se possível estabelecer um *ranking*, de modo a se considerar como “estado-da-arte” o algoritmo que atingir o maior valor na métrica de erro escolhida para comparação.

Na *web* há uma base *online*, frequentemente atualizada, denominada *Papers With Code*, onde são considerados apenas algoritmos publicados em artigos, com código fonte aberto disponível no *gitHub*, para reprodutibilidade. A seção específica do *site*, onde o tópico da Segmentação Semântica é abrangido, está disponível em: <<https://paperswithcode.com/task/semantic-segmentation>>. São apresentados diversas comparações, de acordo com o conjunto de dados público utilizado. A métrica utilizada para comparação é o IoU médio. Aqui se avançará na discussão utilizando os resultados obtidos utilizando o *dataset Cityscapes* (CORDTS, et. al., 2016). Na figura 1.5 está apresentado um gráfico com a evolução semestral do estado-da-arte desde Junho de 2015.

Figura 1.5: Evolução do Estado-da-Arte em Segmentação Semântica



FONTE: <<https://paperswithcode.com>>

Os métodos destacados em vermelho na figura 1.5 serão apresentados, em ordem crescente e de maneira detalhada, na tabela 1.2. Ao leitor, recomenda-se o acesso à versão interativa no *site*, por meio do link apresentado, onde será possível

acompanhar de maneira interativa o *ranking* e observar detalhadamente os códigos e artigos de cada algoritmo do mesmo.

Tabela 1.2: Evolução do Estado-da-Arte em Segmentação Semântica

Algoritmo	IoU médio	Semestre/Ano	Autor(es)	Título
CRF-RNN	62.5	01/2015	ZHENG, et. al.	Conditional Random Fields as Recurrent Neural Networks
RefineNet	73.6	02/2016	LIN, et. al.	RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation
PSPNet	81.2	01/2017	ZHAO, et. al.	Pyramid Scene Parsing Network
HRNetV2 + OCR	83.7	02/2019	YUAN, et. al.	Object-Contextual Representations for Semantic Segmentation

FONTE: Sdaptado de <<https://paperswithcode.com>>

Assim sendo, pode-se afirmar que o estado-da-arte em segmentação semântica, considerando uma classificação multi-classes em cenas terrestres urbanas é de 83,7 % de qualidade, expressa por meio da métrica IoU (Figura 3.13).

1.5 PREÂMBULO E ESTRUTURA DO TRABALHO

Após a formal apresentação das justificativas e motivações por trás do trabalho desenvolvido, já é possível fornecer ao leitor um breve preâmbulo.

O presente trabalho apresenta dois grandes temas ao longo de seu corpo. O primeiro envolve todas as etapas do desenvolvimento de um SMMT de baixo custo com a capacidade de capturar dados espaciais com a vocação primeira de servir ao mapeamento e estudos de vegetação urbana. O segundo é o treinamento e uso de uma rede neural convolucional para classificação das imagens tomadas em vegetação e não-vegetação. O trabalho está organizado em 5 capítulos, obedecendo a seguinte subdivisão:

- No capítulo 1 estão apresentados os conceitos introdutórios fundamentais para apresentação do trabalho ao leitor. Estão inclusos os objetivos gerais e específicos do trabalho, assim como reflexões a respeito das motivações relacionados à escolha da temática e de sua relevância. Ao fim é apresentada uma seção comentando sobre o estado da arte sobre os tópicos mais relevantes no âmbito do trabalho.
- No capítulo 2 está apresentado o arcabouço teórico que alicerça o presente trabalho. No capítulo 3 é apresentada a metodologia, inicialmente com enfoque nos recursos utilizados. Na sequência os procedimentos desenvolvidos e seu emprego são apresentados.
- No capítulo 4 são apresentados os resultados e análises quantitativas e/ou qualitativas dos mesmos.
- No capítulo 5 estão apresentadas as considerações finais do trabalho, assim como uma série de ideias para trabalhos futuros;
- Nos Apêndices, estão apresentados conteúdos extras, objetivando expandir a discussão. Como por exemplo o Apêndice B, onde estão apresentados procedimentos para extração simples de medidas lineares nas fotografias, que podem ser úteis para o emprego de diversas técnicas da área de Engenharia Florestal¹⁶, viabilizando estudos e inferências sobre as árvores urbanas. Um exemplo simples seria uma rápida checagem de se uma árvore visivelmente adulta está de acordo com parâmetros apresentados na tabela 1.1.

¹⁶ Integrando uma ciência denominada Dendrometria, que a partir de grandezas medidas como o diâmetro, busca estimar outras grandezas, como altura, volume e idade das árvores. Fonte: SILVA, J. D., & NETO, F. D. P. (1979). *Princípios básicos de dendrometria*. Recife: Universidade Federal Rural de Pernambuco, Departamento de Ciência Florestal.

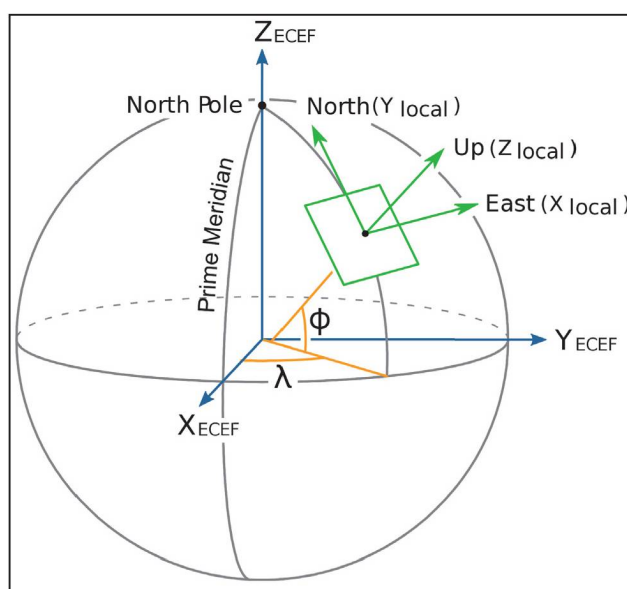
2 FUNDAMENTAÇÃO TEÓRICA

2.1 SISTEMAS DE NAVEGAÇÃO E SISTEMAS DE COORDENADAS

A existência de um sistema de navegação pressupõe a existência de um sistema de coordenadas. De acordo com KALNINS (2009), um sistema de coordenadas tem como finalidade determinar unicamente a posição de qualquer ponto no espaço, podendo ser parametrizado em termos de coordenadas de diferentes tipos: cartesianas, polares, cilíndricas, entre outras.

Nas ciências geodésicas, além dos sistemas de coordenadas em um esferóide oblato (também conhecidas como *coordenadas geodésicas*), é realizado vasto emprego de sistemas de coordenadas cartesianas. Tal referencial trata o espaço como Euclidiano e é dotado de uma tríade de eixos ortogonais (COLLINS, 1989). Os sistemas de coordenadas geodésico, cartesiano geocêntrico (ECEF) e geodésico local (SGL), estão apresentados na figura 2.1.

Figura 2.1: Sistemas Geodésicos de Coordenadas

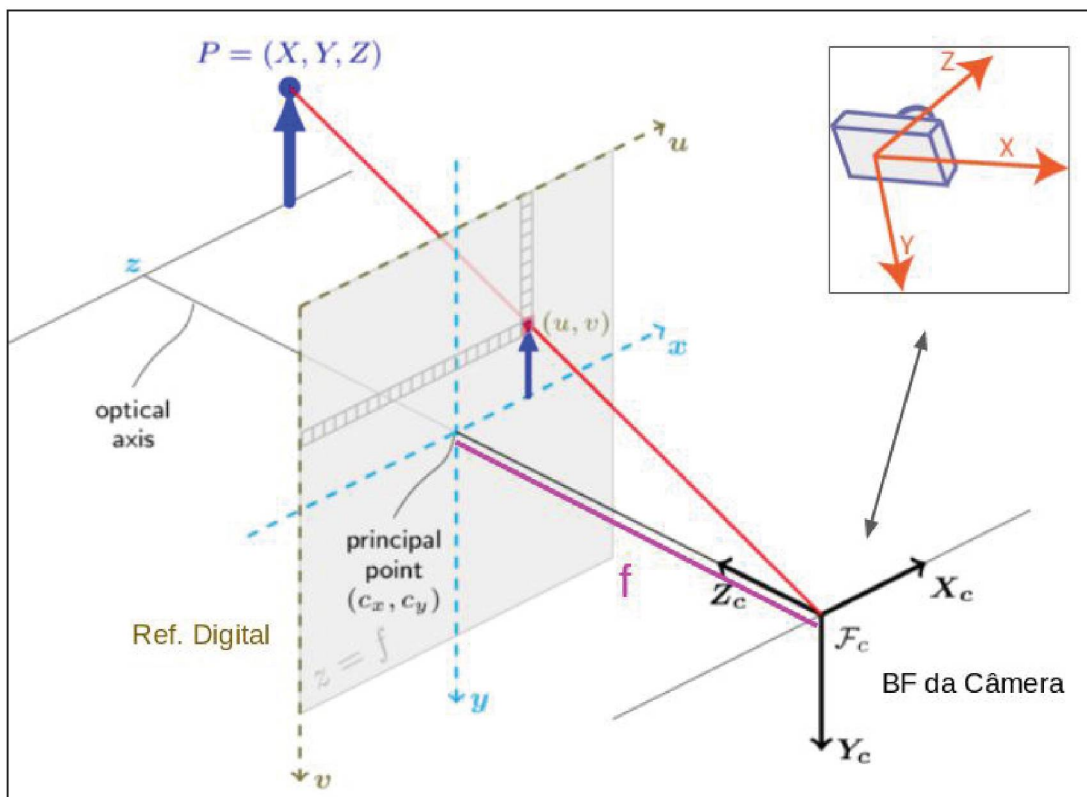


FONTE: MASIULIONIS (2016)

Sobre os sistemas de eixos, é possível definir a orientação de um corpo rígido por meio da atribuição de um sistema de eixos fixo a tal corpo, seu *Body-Frame* (BF). A orientação será definida em função da rotação necessária para transformar

de um sistema de eixos para o outro. Um exemplo de BF (câmera) está demonstrado na figura 2.2, juntamente como referencial digital de um sensor de imageamento.

Figura 2.2: BF de uma câmera e referencial. digital



FONTE: adaptado de <<https://docs.opencv.org>> e <<https://www.mathworks.com>>

2.1.1 Quatérnios Unitários Como Parametrização da Orientação 3D

De acordo com o Teorema de Euler sobre rotações, qualquer sequência de n rotações em três dimensões pode ser substituída por uma rotação única ao redor de um eixo apropriado, posteriormente denominado *Eixo de Euler* (adaptado de Palais & Palais, 2007).

O quatérnio unitário é formado pela soma de um vetor tridimensional com um escalar, suas 4 componentes (x, y, z e w) se interpretadas como um vetor quadridimensional, constituem-no com norma unitária. Os quatérnios podem ser apresentados de acordo com a equação 2.1 (adaptado de DIEBEL, 2006).

$$q = q_w + \vec{q} = f(\theta, \hat{e}) = \cos(\theta) + \hat{e} * \sin(\theta) \quad (2.1)$$

Onde:

- $\sqrt{q_x^2 + q_y^2 + q_z^2 + q_w^2} = 1$
- θ : é o ângulo de rotação tal qual demonstrado no teorema de Euler;
- \hat{e} : é o eixo de Euler, parametrizado como um vetor unitário;

A aplicação de uma rotação parametrizada como um quatérnio pode ser realizada por meio da equação 2.2, sendo p' o vetor rotacionado e p o vetor original, aqui a multiplicação representa a multiplicação de quatérnios¹⁷, que pode também ser empregada para , onde p deve ser expresso como um quatérnio com componente escalar igual a zero.

$$\vec{p'} = q * \vec{p} * q^{-1} \quad (2.2)$$

Onde:

- $q^{-1} = q_w - \vec{q}$

2.2 SISTEMAS DE POSICIONAMENTO GLOBAL POR SATÉLITES

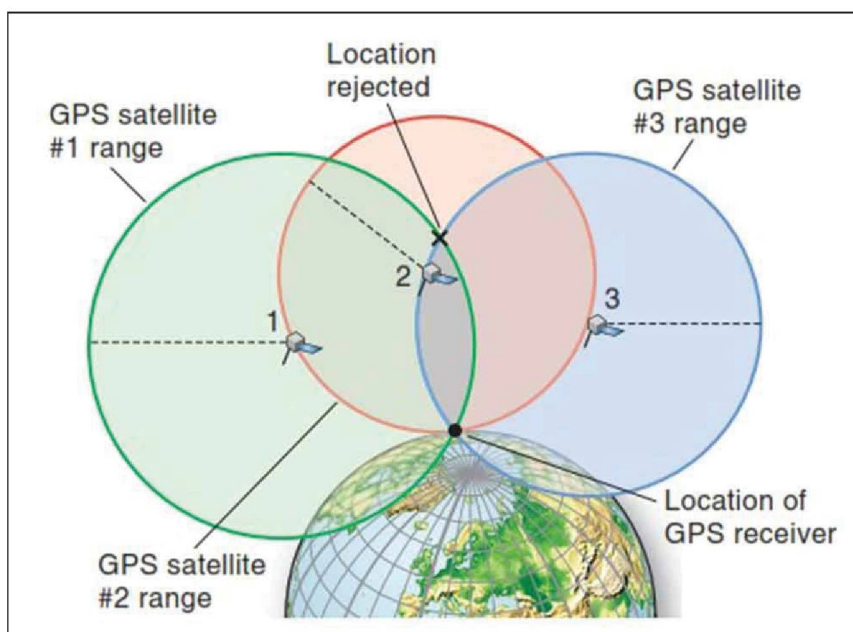
O primeiro sistema de posicionamento global baseado em satélites foi o TRANSIT, baseado no efeito doppler, com um único satélite sendo rastreado a partir de várias posições em seu deslocamento (STANSELL, 1983). Após o TRANSIT, entrou em cena o *GPS - Global Positioning System* (Sistema de Posicionamento Global) e posteriormente outras constelações, como o GLONASS e o GALILEO,

¹⁷ Operação que também pode ser empregada para gerar o quatérnio correspondente combinação da rotação composta por um quatérnio q_1 e então um q_2 . A multiplicação de quatérnios não é comutativa. como qualquer combinação de rotações em espaços euclidianos. Aqui não se entrará em detalhes a respeito da mesma, recomendam-se os artigos disponíveis em <www.sciencedirect.com>.

constituindo o *Sistema de Navegação Global por Satélites - GNSS* (NAVIPEDIA, 2014).

O GNSS tem como idéia básica o uso dos satélites como pontos de referência, “triangulando” a posição a partir do conhecimento de distâncias para três (geometricamente, como demonstrado na figura 2.3) satélites com coordenadas conhecidas. A distância é determinada de acordo com o tempo de viagem necessário para que uma onda eletromagnética emitida por um satélite chegue ao receptor. Para que o posicionamento seja acurado o suficiente, a medida de tempo precisaria estar acurada em ordens de grandeza numericamente muito pequenas, algo que não é possível fazendo uso de osciladores de quartzo.

Figura 2.3: Geometria do Posicionamento por Satélites



FONTE: SHARDA (2018)

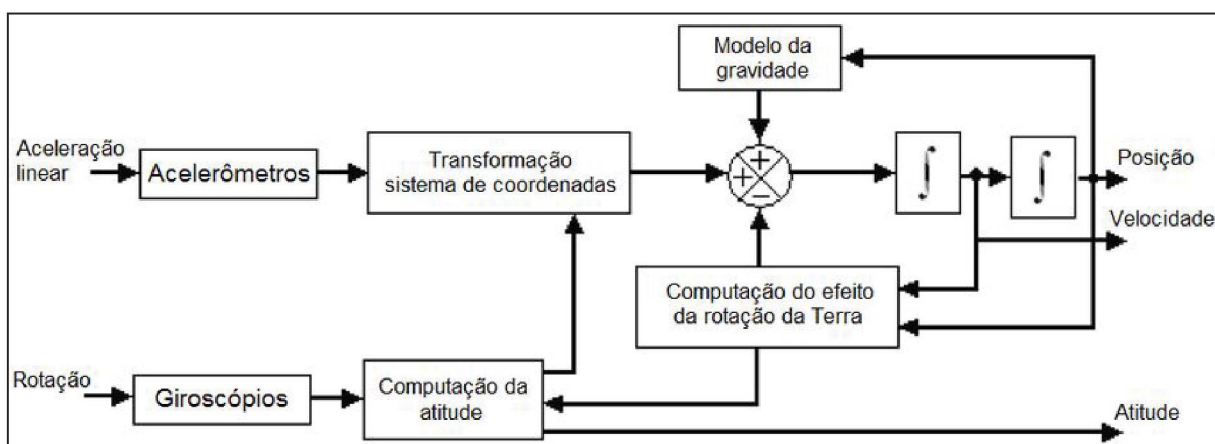
A acurácia do relógio dos satélites é matematicamente “transferida” para o receptor, por meio da inclusão de uma variável para o atraso do relógio do receptor no sistema de equações, o que exige um satélite além do mínimo necessário geometricamente. A acurácia no posicionamento GNSS, passa pela modelagem de erros sistemáticos que afetam sobretudo o tempo de propagação da onda, os principais erros no posicionamento GNSS incluem: erro do relógio do satélite; erro da órbita do satélite; atraso causado pela ionosfera; atraso causado pela troposfera; e multicaminhamento. A mitigação das fontes de erros passa pelo uso de técnicas

como o posicionamento relativo, o uso de várias frequências, de várias constelações e o chamado “Posicionamento por Ponto Preciso”, que inclui a modelagem por meio do fornecimento de modelos baseados em estações de monitoramento contínuo (JEFFREY, 2015).

2.3 SISTEMAS DE NAVEGAÇÃO INERCIAL

Os sistemas de Navegação Inercial - INS, são os únicos capazes de determinar onde um corpo está, esteve e irá, apenas fornecendo a pose inicial, tudo isso por meio das leis da física. Tais sistemas, modernamente funcionam por meio de tríades de acelerômetros e giroscópios ligados a um sistema computacional de controle. Os acelerômetros sentem a aceleração total ao longo da direção com a qual estiverem alinhados, após separar a componente gravitacional (por meio do conhecimento do campo gravitacional local) da aceleração livre, é possível integrar duas vezes essa última de modo a obter a variação da posição. Os giroscópios sentem a velocidade angular total ao redor da direção com as quais estão alinhados, após a eliminação da parcela explicada pela rotação terrestre, é possível integrar a velocidade angular restante de modo a se obter a variação da orientação. É a única técnica que pode ser considerada totalmente autocontida, por não sofrer com distúrbios locais do meio e só depender de seu estado interno (TAZARTES, 2014). O Fluxograma simplificado de funcionamento de um INS está apresentado na figura 2.4.

Figura 2.4: Fluxograma do Funcionamento de um INS



FONTE: (SILVA, et. al., 2014)

As maiores fontes de erros sistemáticos em sistemas de navegação inercial, são: leitura diferente de 0 na ausência de sinal (*bias offset*); a presença de um fator de escala; a não-ortogonalidade dos eixos dos sensores; sinal anômalo causado por aumento da temperatura; e o ruído do sensor. O ruído é classificado como “branco” e gaussiano, podendo ser abordado estocasticamente; os efeitos causados pela temperatura exigem que o sistema seja mantido em temperatura constante ou que as mudanças de temperatura sejam medidas de modo a propiciar que o erro seja matematicamente mitigado; os demais erros exigem procedimentos de calibração (WOODMAN, 2007).

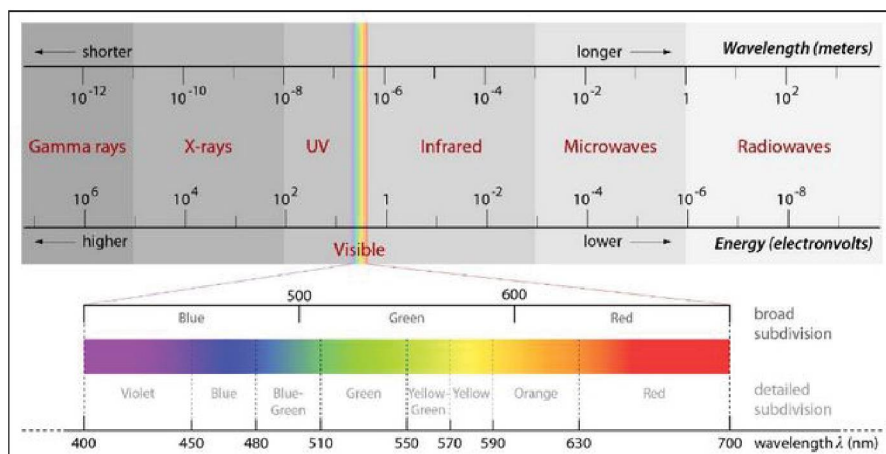
2.4 O ESPECTRO ELETROMAGNÉTICO E SENSORES PASSIVOS

Uma das formas de a energia viajar no espaço é por meio de ondas eletromagnéticas, tais ondas são uma forma de radiação caracterizada pelo seu comprimento (inv. proporcional à sua frequência) e por sua intensidade (amplitude). O olho humano é sensível a parte de tal espectro, que geralmente é chamado de “luz”. Há de fato uma dualidade onda-partícula, de modo que as formas de “luz” podem ser quantificadas de maneira discreta. Tais partículas, denominadas “fótons”, interagem com a matéria, sendo que quando esta é exposta, alguns fótons são absorvidos, outros refletidos/refratados e alguns eventualmente atravessam os corpos. Para cada combinação de material, geometria de incidência e comprimento de onda, haverá uma combinação diferente de interação (adaptado de KHAN ACADEMY, 2019). O espectro eletromagnético está esquematizado na figura 2.5.

Sensores digitais como o CCD e CMOS são capazes de transformar os fótons (originários geralmente da porção refletida da luz solar após a interação da mesma com os objetos no ambiente) que incidem sobre seus elementos sensores, gerando uma corrente elétrica. Uma vez que trata-se de uma matriz de elementos sensores, é gerada uma imagem bidimensional em perspectiva da cena em questão, sendo que a luz é direcionada para a região do sensor por meio de um sistema de lentes, para que haja luz suficiente, uma vez que a luz passará por uma pequena abertura, para que a imagem se forme de maneira nítida (Adaptado de LUCID VISION LABS.,

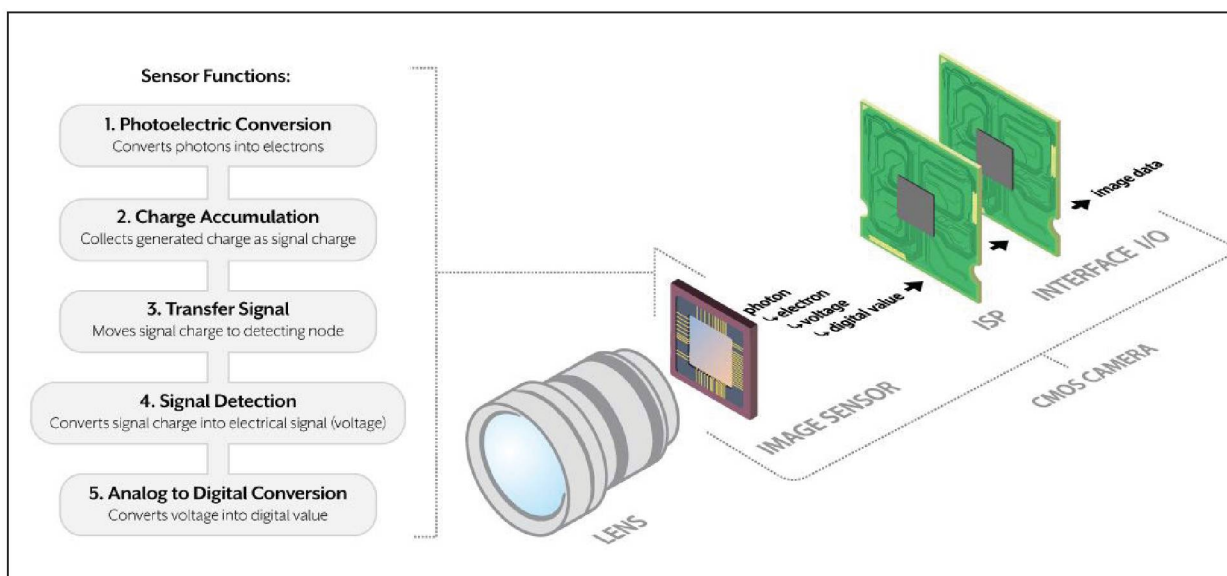
2019). Na figura 2.6 está esquematizado o processo de transformação da luz em imagens digitais.

Figura 2.5: O Espectro Eletromagnético



FONTE: VERHOEVEN (2017)

Figura 2.6: Fluxograma de transformação de Luz em Imagens Digitais



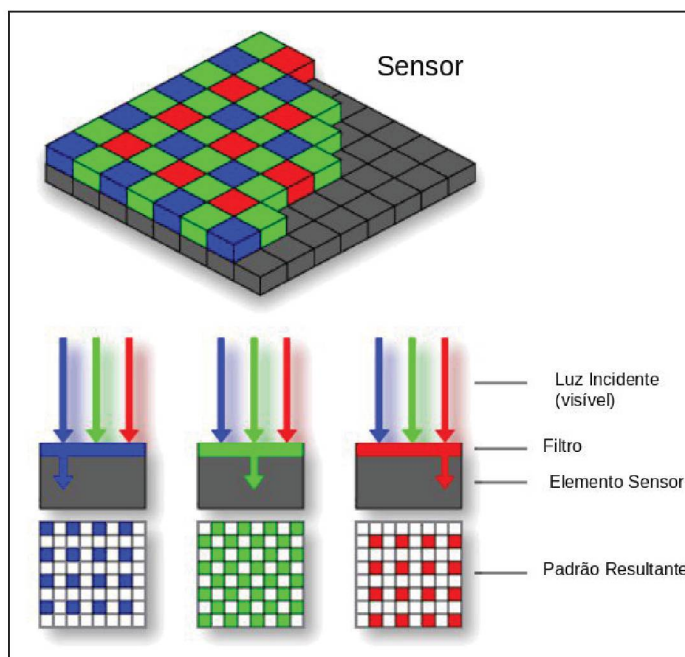
FONTE: <<https://thinklucid.com/tech-briefs>>

2.4.1 Imageamento R-NIR com Aplicação de Filtro Vermelho

A sensibilidade espectral dos sensores digitais é diferente da humana. Tais sensores (CCD/CMOS) são capazes de capturar luz em frequências mais baixas que o Violeta e mais altas que o Vermelho. Entretanto, o sensor por si só não é

capaz de distinguir as cores, uma vez que a energia chega como um todo para o mesmo. Assim se faz necessário o emprego de filtros, compostos por materiais translúcidos, mas apenas para parte do espectro eletromagnético. Para imagens normalmente se emprega a combinação de 2 filtros: o primeiro é um filtro que só permite a passagem da luz visível, afetando o sensor inteiro¹⁸; sobre todos os pixels é inserido o chamado *filtro de bayer*, que para cada pixel só permite que uma das 3 bandas R, G ou B passe (adaptado de MCHUGH, 2017). O filtro de Bayer está demonstrado na figura 2.7.

Figura 2.7: O Filtro de Bayer

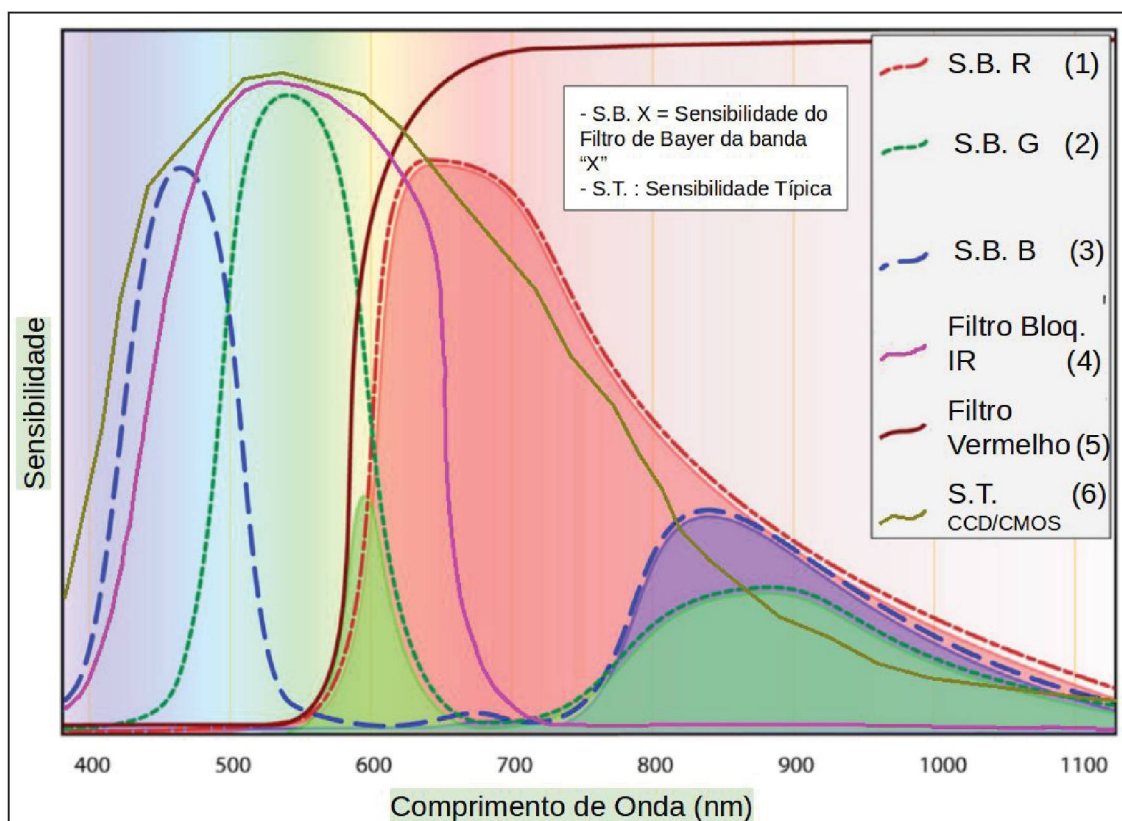


FONTE: Adaptado de SIORDIA, et al. (2011)

Caso o filtro para luz não-visível seja removido, será possível explorar toda a extensão da sensibilidade espectral do sensor utilizado. Contudo, pensando em aplicações específicas, é possível inserir outros filtros, que irão alterar as cores da imagem resultante. Entretanto geralmente não é possível remover o filtro de bayer. Na figura 2.8 estão apresentadas uma série de curvas de Sensibilidades Espectrais Típicas para demonstrar o efeito da aplicação de um filtro vermelho genérico em frente a um típico sensor digital.

¹⁸ Geralmente é inserido próximo às lentes ou em cima do sensor.

Figura 2.8: Curvas de Sensibilidade Espectral envolvendo sensores e filtros



FONTE: ADAPTADO DE NIJLAND, et. al. (2014) e <www.ir-photo.net>

Pode-se observar na imagem que o filtro de bloqueio de IR(4), restringe a captura basicamente apenas ao espectro do visível, sendo que cada camada¹⁹ (1-3) do filtro de Bayer possui uma sensibilidade distinta, sendo que muitas frequências são capturadas por mais de uma banda. A remoção do filtro 4 permite que o sensor (6) capture toda a faixa à qual é sensível. A inserção de um filtro de barreira Vermelho (5) impede que as frequências maiores que tal cor sejam captadas (adaptado de NIJLAND et. al., 2014). Assim, após a inserção de tal filtro, como visível na imagem, praticamente apenas informação da região do infravermelho será capturada pela banda B; na banda R haverá um grande “vazamento” de Infravermelho, entretanto o pico de sensibilidade das frequências do vermelho é muito maior; e na banda G haverá a maior mistura.

¹⁹ Aqui o termo não faz alusão a uma idéia de “sobreposição”.

2.4.2 NDVI

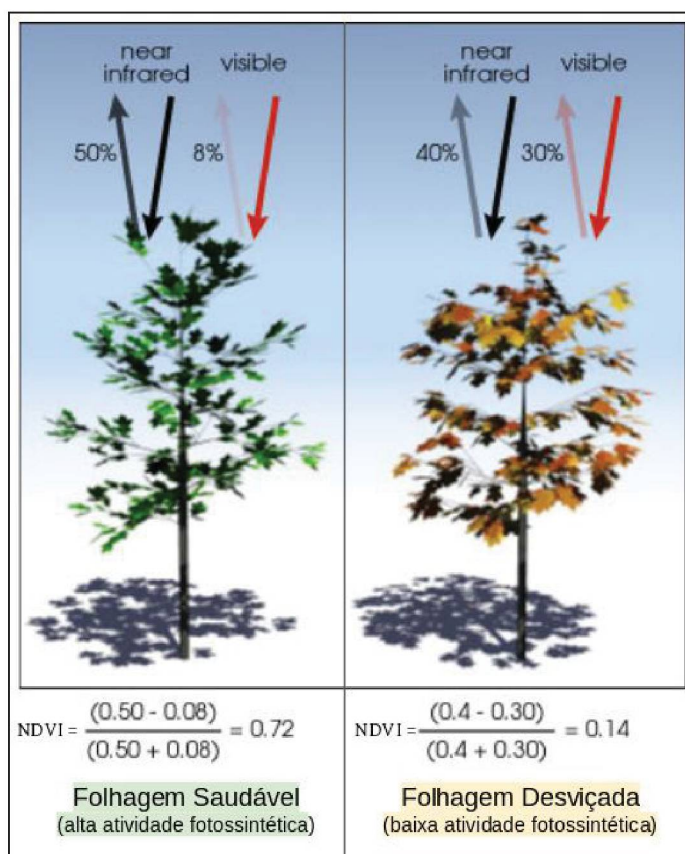
Diferentes imageadores observam diferentes partes do espectro eletromagnético e uma vez que a luz solar se apresenta muito além do visível (tanto em frequências mais baixas que o violeta, quanto mais altas que o vermelho) é possível capturar a resposta espectral dos mais diversos alvos, geralmente sobre a superfície terrestre. Com relação à vegetação, o pigmento presente nas folhas (clorofila) absorve intensamente a luz visível, em contrapartida, a estrutura foliar reflete em grande escala a parte correspondente ao infravermelho próximo. Na década de 1970, cientistas descobriram que a absorção de luz solar de uma planta é diretamente proporcional à quantidade de fotossíntese que a mesma está realizando.

Assim, do uso da diferença da resposta nas bandas R e NIR, empregada para medir a capacidade fotossintética das folhagens, juntamente com uma normalização empregando a soma dos valores de tais bandas, foi criado o NDVI, apresentado na equação 2.1 (WEIER & HERRING, 2000).

$$NDVI = \frac{NIR - R}{NIR + R} \quad (2.1)$$

Na figura 2.9 estão exemplificadas as respostas espectrais e valores de NDVI para folhagens com diferentes níveis de atividade fotossintética.

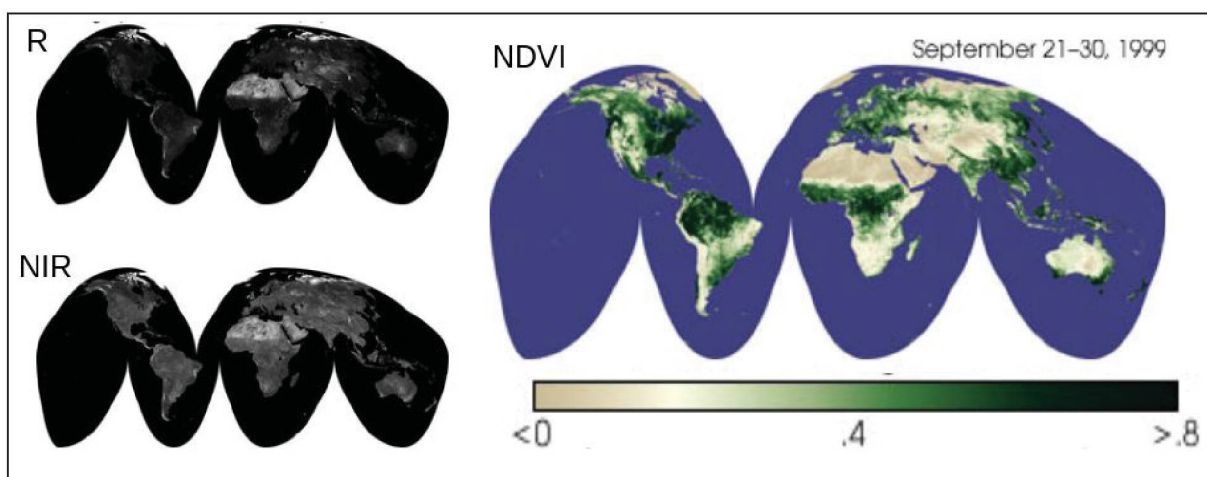
Figura 2.9: Exemplo de respostas espectrais e valores de NDVI para folhagens



FONTE: ADAPTADO DE WEIER & HERRING (2000)

Na figura 2.10 está demonstrado um exemplo de cálculo de valores de NDVI sobre toda a superfície Terrestre para uma data fixa.

Figura 2.10: NDVI Calculado para toda a superfície terrestre



FONTE: ADAPTADO DE WEIER & HERRING (2000)

2.5 SISTEMAS DE MAPEAMENTO MÓVEL TERRESTRE

Um SMMT pode ser definido como uma plataforma móvel onde um Sistema de Georreferenciamento Direto e Sensores Remotos são fixados para a aquisição de dados sincronizados e alinhados no domínio do tempo para determinar a posição tridimensional georreferenciada das feições imageadas pelos sensores. Tal sistema também, contará com todas as ferramentas para armazenar, manipular, processar, classificar e atualizar os dados brutos e algumas informações já processadas. Ainda é colocado como alternativa que esse sistema pode ser auxiliado por odômetros, inclinômetros, bússolas digitais e outros dispositivos (MADEIRA et. al., 2012).

A GEOVAN, que é um SMMT do Instituto Cartográfico da Catalunha, tem um sistema de navegação GPS/INS com duas antenas GPS, duas câmeras em configuração estereoscópica e um laser-scanner móvel. O sistema, é composto pelos seguintes subsistemas: orientação, para defini-la juntamente com a posição; plataforma de integração: para abarcar os sensores; Imageamento, que contém as câmeras e seu sistema de disparo e armazenamento; sincronização, que contém os elementos eletrônicos e computacionais para o alinhamento temporal no tempo GPS dos dados; e o módulo de controle de energia e comando, que contém a alimentação dos equipamentos e o computador responsável pelo comando central do SMMT em questão (Talaya et. al., 2004). A GEOVAN está apresentada na figura 2.11.

Pode-se observar, a partir do que foi apresentado anteriormente em MADEIRA et. al. (2012), que um SMMT deve conter como partes básicas, um sistema de navegação que permita a determinação e orientação global; um sistema imageador que permita coletar feições do ambiente; uma plataforma rígida que não permita que as orientações relativas dos sensores se alterem; uma unidade computacional para armazenamento e processamento de dados; e por fim, uma unidade de alimentação.

Figura 2.11: GEOVAN, o SMMT do ICC

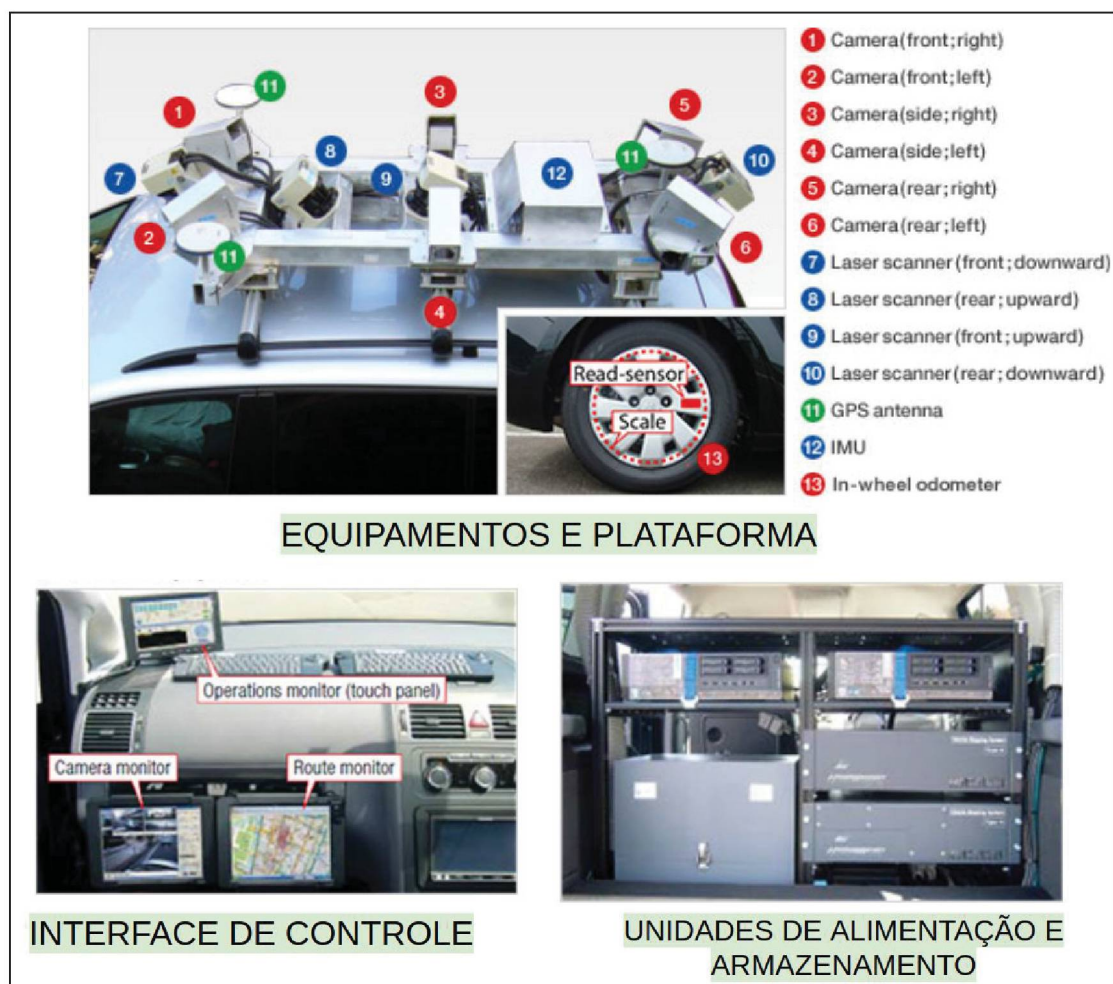


Fonte: Talaya, et. al (2004)

Um SMMT consolidado, é o criado pela Mitsubishi Electric, que afirma apresentar uma acurácia relativa de 1 cm e absoluta de 10 cm. O sistema possui 3 antenas GNSS, uma IMU, 4 câmeras e 4 laser scanners, além de um odômetro de roda²⁰. O SMMT da Mitsubishi está apresentado na figura 2.12.

²⁰ Fonte: <<http://www.mitsubishielectric.com/bu/mms/features/index.html>>.

Figura 2.12: O SMMT Comercial da Mitsubishi



Fonte: <<http://www.mitsubishielectric.com>>

2.6 REDES NEURAIS CONVOLUCIONAIS E *DEEP LEARNING*

Deep Learning é um termo utilizado para um processo que gera modelos (redes neurais) que evoluem suas capacidades, sendo treinados utilizando dados de entrada pouco estruturados e nenhuma instrução explícita, nem conhecimento prévio a respeito de como solucionar o problema. As redes neurais, de forma geral são compostas por *camadas*, contendo várias formas de interagir com a informação e diferentes maneiras de estarem interligadas, sendo que sempre a primeira camada será denominada “camada de entrada” e a última “camada de saída”. Há diversos tipos de redes neurais: redes neurais clássicas; redes neurais recorrentes; redes neurais convolutivas ou convolucionais, etc. (adaptado de DENG & YU, 2014).

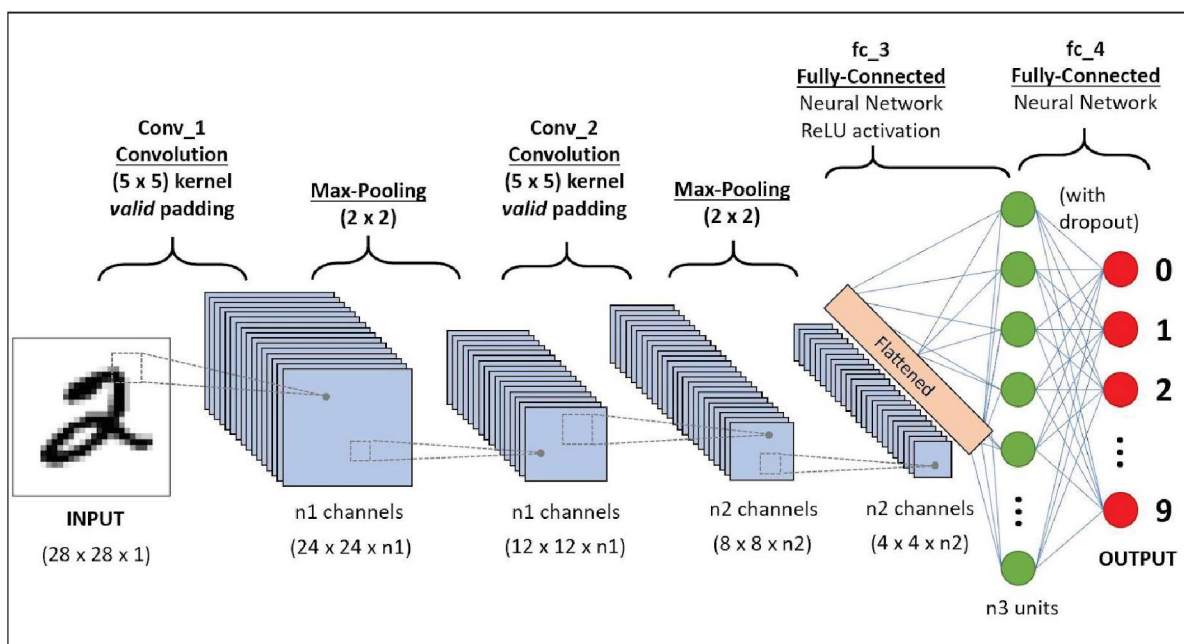
As Redes Neurais Convolutivas (CNN) são redes neurais artificiais projetadas para trabalhar com imagens digitais, isto é: receber uma imagem como entrada e após sua completa passagem pela rede (processo conhecido como *forward pass*) produzir uma saída. Sua unidade básica são filtros baseados em convoluções, tais filtros também são um dos tipos possíveis de camadas das CNNs. Os tipos de camadas mais comuns são apresentados a seguir (adaptado de KARPATY et. al, 2020):

- De Entrada: as imagens fornecidas para a rede, cuja dimensionalidade afetará a dimensionalidade das demais camadas da rede;
- Convolucionais: conectam-se localmente à camada anterior, por meio de máscaras de convolução, também conhecidas como filtros. Convolução é a operação de um produto interno²¹ entre o filtro e uma região na imagem de mesmo tamanho;
- De Ativação: Filtram os valores retornados na camada anterior de modo a realizar uma transferência dos valores retornados para valores mais próximos a 0 ou a 1, funções comumente utilizadas são: sigmóide; tangente hiperbólica; e a unidade retificadora linear (ReLU);
- Agregadoras: atuam realizando a redução de dimensionalidade da camada de entrada, geralmente se usa a “votação pelo máximo 2 x 2”, isto é, para um conjunto de 2 x 2 células, irá continuar apenas o valor máximo. Entretanto podem também ser utilizadas outras funções como a média.
- Completamente Conectadas: a última camada é transformado em um vetor escalares, de modo a criar uma rede neural clássica. Tal tipo camada é geralmente utilizada para produzir saídas textuais;
- De Saída: pode ser outra imagem ou um único número, é basicamente o resultado produzido pela rede;

Na figura 2.13 está esquematizado o *forward pass* de uma CNN para classificação de imagens contendo algarismos indo-arábicos de 0 a 9.

²¹ Generalização n-dimensional do produto escalar: somatório dos produtos dos elementos de mesmo índice.

Figura 2.13: Esquematisação da estrutura de uma CNN simples



Fonte: <<https://towardsdatascience.com>>

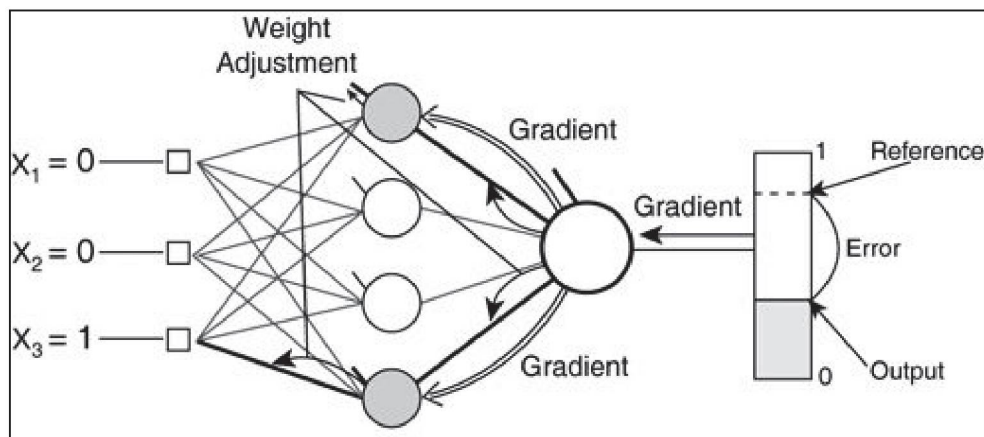
Outra questão é a forma como as camadas são interligadas entre si, tal conceito é conhecido como *frontend*, alguns exemplos serão aqui apresentados:

- **Básico**: uma camada está ligada apenas às suas vizinhas;
- **Resnet**: realiza ligações diretas entre camadas não vizinhas, apresentado em (HE, et. al., 2016);
- **Inception**: combina o resultados de núcleos convolutivos com várias dimensões, apresentado em (SZEGEDY, et. al. 2015);

O treinamento de uma CNN resulta na evolução dos pesos da rede (por exemplo, os valores das máscaras de convolução) e consequentemente na mudança da qualidade das saídas geradas. Tais pesos podem ser inicializados por meio de valores aleatórios, que serão empregados na primeira passagem das imagens de entrada pela rede. Após isso, será calculada uma função-custo que terá como entrada a verificação da diferença (gradiente) das saídas produzidas pela rede com dados de controle (validação). Iniciando pela última camada da rede, a função de erro será combinada com os pesos antigos para calcular os novos pesos, repetindo esse processo para cada camada (por meio da aplicação da regra da cadeia), até chegar na ligação entre a camada de entrada e a primeira camada da

rede. Tal processo é denominado retropropagação (tradução livre), sendo o responsável pela “evolução” da rede (adaptado de DENG & YU, 2014). Na figura 2.14 está ilustrado o processo de retropropagação.

Figura 2.14: Ilustração do processo de Retropropagação

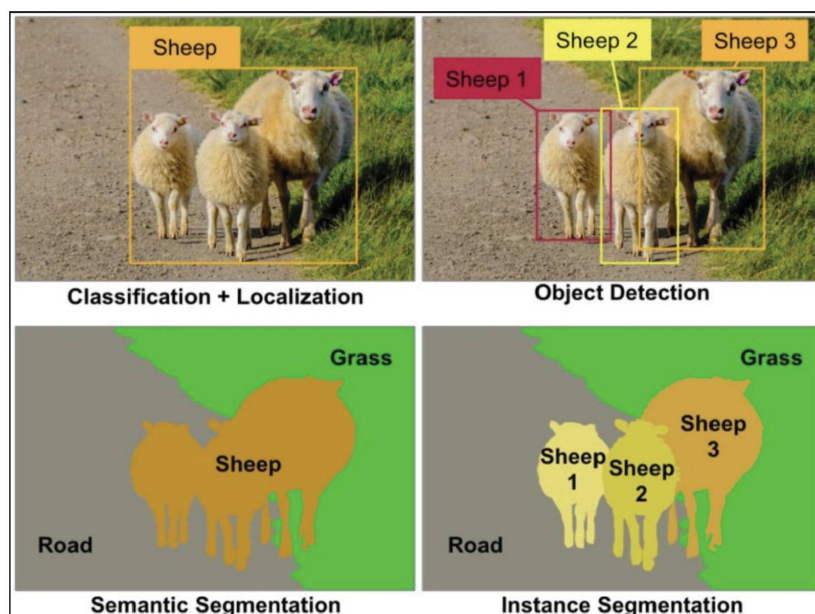


FONTE: <<http://www.yaldex.com/game-development>>

As CNNs podem ser utilizadas para uma série de aplicações. No âmbito da classificação de imagens, as aplicações genéricas se dividem nas seguintes categorias (adaptado de Yin, 2019, traduções livres), demonstradas graficamente na figura 2.15:

- A. Classificação da Imagem: retorna a principal classe presente na imagem. Opcionalmente pode englobar a localização de tal classe, retornando também o retângulo envolvente do objeto presente na imagem;
- B. Detecção de Objetos: detecta eventuais n presenças de n objetos, retornando o retângulo envolvente e o nome da classe a que cada detecção pertence;
- C. Segmentação Semântica: para cada pixel da imagem é atribuída uma classe, de modo a retornar uma imagem com a mesma resolução da entrada, preenchida com as cores correspondentes a cada classe;
- D. Segmentação Semântica e de Instâncias: semelhante ao item anterior, mas caso haja feições individualizáveis, para cada aparição, será inserido um código diferente (decodificado como uma cor diferente) na imagem de saída;

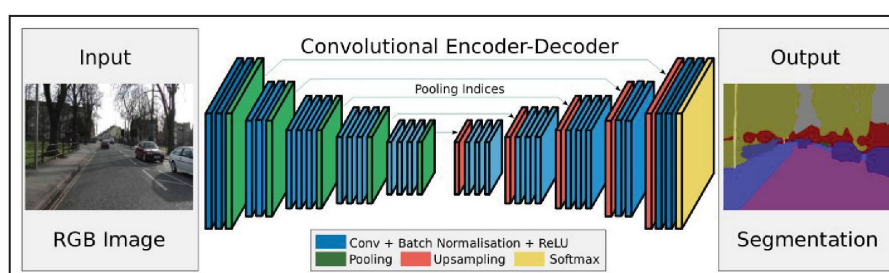
Figura 2.15: Categorias de aplicações genéricas de CNNs em classificação



FONTE: <https://towardsdatascience.com>

Há infinitas possíveis arquiteturas para CNNs. Enfatizando-se as redes voltadas para segmentação semântica, há duas diferenças básicas: geralmente a saída não é produzida por uma camada completamente conectada e são comuns arquiteturas puramente convolucionais, uma estrutura típica é apresentada na figura 2.16. Seus blocos construtivos principais são as camadas convolutivas chamadas de “Codificadores” e de “Decodificadores”. Os primeiros “codificam” as imagens na forma de “mapas de feições”, ao passo que as últimas, recebem saídas como essas e reconstroem imagens que progressivamente irão recuperando a resolução original, até que na última camada, seja utilizada uma função de ativação *softmax*, que transformará os “mapas de feições” na classificação resultante da rede (adaptado de BADRINARAYANAN, et. al, 2017)

Figura 2.16: Exemplo da Estrutura Básica de uma Rede Completamente Convolucional



FONTE: BADRINARAYANAN et. al. (2017)

3 MATERIAIS E MÉTODOS

3.1 MATERIAIS E RECURSOS

Para a condução da presente pesquisa, foram empregados uma série de equipamentos, assim como recursos, sejam estes computacionais, físicos ou procedimentais. Abaixo estão listados os equipamentos utilizados, por ordem de importância. Também são apresentadas sua finalidade no sistema, assim como os motivos que levaram à sua escolha:

- I. Uma Câmera “Raspicam V2” sem filtro para infravermelho. Baseada em um sensor CCD Sony IMX219 com uma resolução máxima de 3280 x 2464 pixels e uma diagonal de 4.6 mm. Foi escolhido devido ao fato de capturar o infravermelho próximo, assim, suas imagens podem ser usadas para estudos de vegetação. Outra razão para a sua escolha foi devido ao fato de o computador da plataforma ter um plug projetado especificamente para ela, de modo que a velocidade de transmissão de dados é otimizada.
- II. Uma módulo *Adafruit* BNO055 com 3 acelerômetros, 3 giroscópios e 3 magnetômetros: utilizados como a unidade de medida inercial para o sistema. Este sensor foi escolhido devido ao custo acessível e ao nível de acurácia esperado.
- III. uma Módulo GNSS-UBX-M8030 com antena e receptor GNSS: servindo como módulo GNSS para o sistema de navegação. Escolhido devido a disponibilidade no mercado nacional e ao custo acessível.
- IV. Um computador Raspberry Pi 3: utilizado para funcionamento simultâneo dos equipamentos do sistema por meio das interfaces para câmera, USB (GNSS) e GPIO (para a IMU), assim como armazenamento e transferência dos dados coletados. Escolhido devido ao seu baixo custo e disponibilidade no mercado nacional.

Pode-se perceber que a escolha dos equipamentos está ligada ao baixo custo e à sua disponibilidade, uma vez que o protótipo a ser desenvolvido foi projetado

para a maior reprodutibilidade possível, além de ser acessível, com a restrição ligada exclusivamente ao conhecimento técnico.

Adicionalmente aos elementos essenciais para a composição do sistema de navegação, também são necessários equipamentos adicionais para funções auxiliares, a saber:

- I. Plataforma feita em MDF, para embarcar, de maneira rígida (de modo a não permitir movimentos relativos entre os sensores), os equipamentos. A plataforma foi pintada de branco para reduzir a absorção de umidade e de calor²².
- II. Película da marca Rosco, modelo *Fire*, responsável por bloquear a passagem das faixas do espectro correspondentes ao azul e ao verde;
- III. Cartão micro SDXC de 32 GB, classe U3²³
- IV. Bateria de 12V com capacidade de 7Ah, feita de chumbo-ácido, para alimentação dos equipamentos da plataforma;
- V. Conversor de voltagem, tendo 14V a 12V como entrada e 5V como saída, para adequar a voltagem da bateria com a voltagem requerida pelo computador da plataforma, que por sua vez é responsável por alimentar os demais componentes;
- VI. Cabos para conexão dos componentes: cabos USB de ligação das câmeras ao computador; cabo USB para ligação da antena GNSS ao computador; cabos para conexão da IMU ao computador; cabos para conexão da bateria ao conversor de voltagem; cabo para conexão do conversor de voltagem ao computador;
- VII. Subplataforma para abrigar a IMU, incluindo proteção térmica (XI) da mesma. Tal item pode ser desacoplado da plataforma de modo a permitir movimentação independente;
- VIII. Um “tapa-sol” para a câmera, de modo a reduzir a ocorrência de *glare*²⁴ nas imagens, devido a incidência direta de raios do sol para as lentes;

²² Fatores que podem influenciar a qualidade do posicionamento.

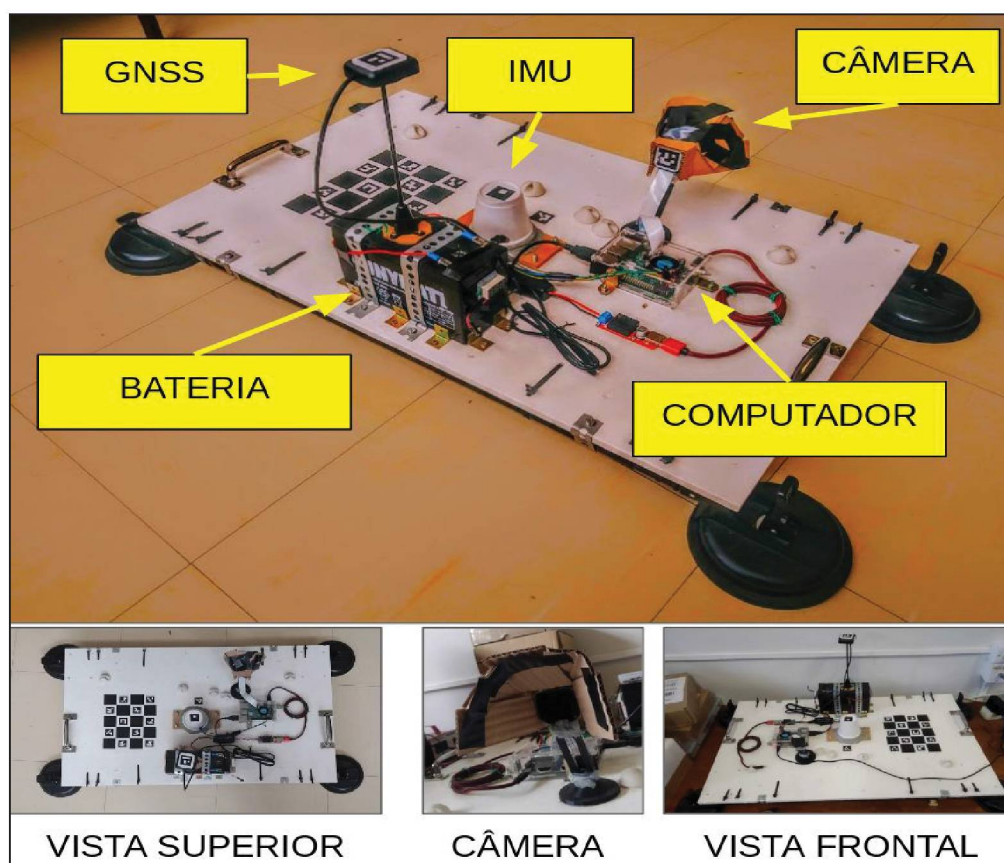
²³ com até 100 MB/s de velocidade de gravação, de acordo com o fabricante: <https://www.sandisk.com.br/home/memory-cards/microsd-cards/extremepro-microsd-uhs-ii>

²⁴ Ponto de brilho excessiva na imagem, causado por incidência direta dos raios solares.

- IX. Interruptor para ligar/desligar os equipamentos da plataforma;
- X. Duas barras de fixação removível da plataforma a um veículo terrestre, por meio de ventosas²⁵;
- XI. Proteção térmica em isopor na forma de tronco de cone para a IMU²⁶;
- XII. Duas alças para transporte da plataforma;
- XIII. Itens para conferir firmeza e estabilidade: parafusos e porcas; abraçadeiras metálicas ou de nylon.
- XIV. Organizadores de cabos, para propiciar maior segurança;

Na figura 3.1 é apresentado a plataforma do SMMT desenvolvida neste trabalho.

Figura 3.1 O Protótipo do SMMT Proposto e seus elementos



FONTE: O AUTOR (2019)

²⁵ de modo a ser possível acoplar o sistema a qualquer veículo, sem que o mesmo possua uma estrutura de fixação, tornando o sistema mais universalmente empregável.

²⁶ um dos fatores de estabilidade das medições da IMU é a manutenção da temperatura.

Com relação aos custos da plataforma, estão apresentados os custos individuais de cada item na tabela 3.1.

Tabela 3.1 Custos individuais dos equipamentos do protótipo do SMMT

Equipamento	Quantidade	Valor Un. (R\$)	Subtotal (R\$)
Câmera Pi-NOIR 8MP	1	180,00	180,00
IMU ada-BNO055	1	175,90	175,90
GNSS-UBX-M8	1	207,41	207,41
Raspberry Pi 3	1	194,99	194,99
Rack Ventosas	1	138,90	138,90
Bateria 12v 7Ah	1	59,00	59,00
Cartão	1	87,99	87,99
Plataforma e Acessórios	1	150,00	150,00
Total (R\$)			1194,19

FONTE: O AUTOR (2019)

Como recursos computacionais, para operacionalização do sistema, foram empregados os seguintes itens:

- I. Sistemas operacionais:
 - A. Ubuntu 16.04 armhf no computador embarcado;
 - B. Ubuntu 18.04/16.04/14.04 desktop em estações computacionais para processamento offline dos dados coletados pela plataforma;
- II. *Robot Operating System* (ROS), *framework* que possibilita a integração e intercomunicação de sensores, além de pré processamento de dados e gravação dos mesmos em um mesmo sistema de tempo.
- III. O protocolo VNC, acessado por meio da interface implementada no *software* RealVNC, para acesso remoto do computador da plataforma;
- IV. Linguagens de programação:
 - A. Python
 - B. *bash scripting*

- V. A biblioteca/framework Google *Tensorflow* para realizar o treinamento e a aplicação das redes neurais empregadas para a segmentação das imagens.
Bibliotecas auxiliares:
 - A. Bibliotecas estatísticas (*numpy*, *matplotlib*, *pandas*) para o cômputo de estatísticas descritivas dos resultados;
- VI. A plataforma de computação *Google Colab*, que permite o uso de máquinas remotas com capacidade de processamento superior.
- VII. O protocolo GIT e o sistema *Github*, para gerenciamento de versões e armazenamento do código em um servidor remoto, de modo a permitir acesso e edição de qualquer localidade.

Todos os frameworks e códigos utilizados trabalham de modo integrado ao ROS, por mais que a plataforma do presente trabalho não desempenhe funções robóticas²⁷, possui sistemas sensores que seriam essenciais para um robô, de modo que um framework voltado à robótica, possui várias das funcionalidades demandadas pela plataforma.

Todos os códigos desenvolvidos para a realização do presente trabalho estão disponíveis em: < <https://github.com/kauevestena/snav> >, organizados em um repositório no *GitHub*. Os códigos seguem uma licença de código aberto, não havendo restrições para seu uso e modificação, exceto reconhecimento de sua origem.

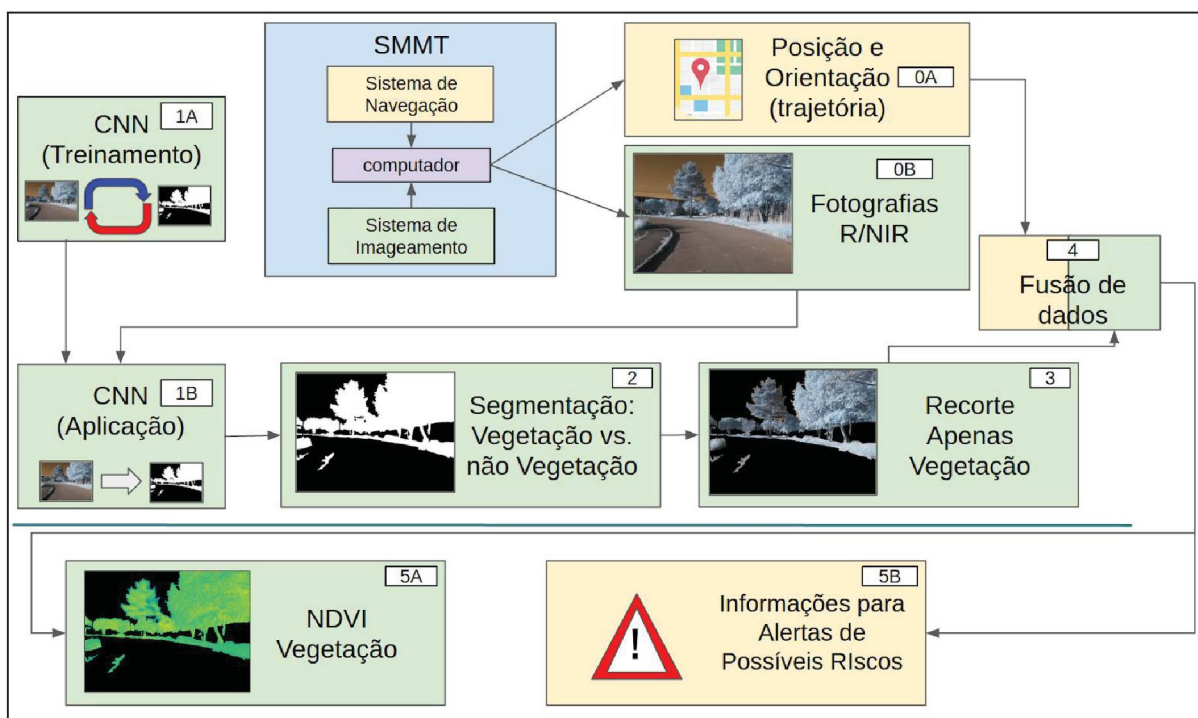
3.2 MÉTODO

3.2.1 Visão Geral Do Método

A presente seção tem como objetivo apresentar o fluxograma de etapas do método desenvolvido neste trabalho. Tal fluxograma está apresentado na figura 3.2 em sua versão simplificada.

²⁷ Leia-se “agir de maneira autônoma”.

Figura 3.2: Fluxograma das etapas do método



FONTES: O AUTOR (2019)

Como mostra a Figura 3.1, o SMMT é a estrutura física responsável pela captura de todos os dados, sendo acoplado em um veículo terrestre, que pode percorrer as vias transitáveis para fins de imageamento do ambiente. Os dados coletados pelos sensores embarcados na plataforma terrestre são brevemente apresentados, considerando dois principais sub-sistemas:

- Sistema de Navegação:
 - Orientação no espaço (IMU)
 - Posição 3D geolocalizada (Antena/Receptor GNSS)
 - Aceleração livre em um referencial global (IMU)
- Sistema de Imageamento:
 - Imagens R/NIR²⁸ em resolução de até 8MP (Câmera)

²⁸ O Sensor utilizado captura, além do visível, o espectro do Infravermelho Próximo. Devido a um filtro instalado na parte frontal das lentes, os sinais correspondentes ao espectro do Azul e do Verde não chegam ao sensor, dentro das limitações do filtro.

O instante da aquisição de todos os dados é registrado com relação a um sistema único de tempo, graças ao computador, ao qual os sensores estão conectados. Acoplado à plataforma, tal computador também é responsável pelo armazenamento destes dados capturados. Após um processo de interpolação linear, efetua-se, matematicamente, a sincronia dos dados de navegação com as imagens.

Para o emprego das imagens tomadas pelo sistema (etapa 0 na figura 3.2), se faz necessário classificar apenas o que se trata de vegetação (etapa 2). Como já mencionado no capítulo 1, diversos outros detalhes pertencentes a uma cena urbana estão presentes nas imagens, assim, sua eliminação irá favorecer etapas posteriores de análise, no âmbito de estudos da vegetação urbana.

A referida classificação será realizada por meio da aplicação de uma CNN (etapa 1B), cujo treinamento (etapa 1A) é realizado por meio de um algoritmo de *Deep Learning*, que recebe como entrada pares de imagens e classificações realizadas manualmente. Iterativamente a qualidade da classificação irá evoluindo, até um patamar máximo, detectado por meio de testes em amostras de controle externo.

Após a extração de um recorte “apenas vegetação” (etapa 3) utilizando a máscara de segmentação e a imagem original, poderão ser realizadas as mais diversas análises com combinações envolvendo os dados do SMMT (“etapa” 4).

O primeiro produto para análise produzido é a imagem contendo os valores calculados de NDVI para a parte correspondente a vegetação(etapa 5A).

A partir do recorte de vegetação e da imagem de NDVI²⁹ seria possível proceder com análises capazes de inferir sobre a saúde da vegetação³⁰. Entretanto, vale ressaltar que as análises necessárias para tal tipo de inferência são complexas e estão além do escopo do presente trabalho, não tendo sido desenvolvidas.

²⁹ Assim como de outros índices possíveis de serem calculados com as bandas R e NIR.

³⁰ Uma vez que um problema fosse detectado, os dados de posição e orientação do sistema de posicionamento e navegação, iriam possibilitar a geração de alertas geolocalizados.

3.2.2 Aquisição dos dados

Neste subitem, serão abordados os procedimentos para obtenção dos dados de interesse, estando divididos em:

1. Captura: dedica-se a apresentar a forma como os dados são gravados pelo SMMT, independentemente da finalidade, ou seja: como são gerados os arquivos de dados.
2. Coleta: aplica-se aos procedimentos de campo para obtenção de dados a serem empregados em processamento posterior.

3.2.2.1 Captura de Dados

Inicialmente, cabe ressaltar que os dados são gravados para pós-processamento, ainda que seja possível se imaginar um cenário onde os procedimentos de processamento sejam realizados em tempo real, tal possibilidade não foi levada em consideração no presente trabalho.

Todos os dados coletados ficam gravados no cartão SD do computador embarcado no sistema, podendo ser acessados posteriormente por meio de uma porta USB do mesmo.

O computador embarcado na plataforma se trata de um computador completo, possuindo uma interface *desktop* como qualquer outro. Os equipamentos da plataforma estão integrados por meio do *framework* ROS Kinetic, de modo a estarem co-registrados com relação ao tempo³¹. Assim para a gravação de dados deve-se demandar três tarefas básicas para o ROS, como segue:

1. Inicializar os *drivers*³² dos equipamentos
 - a. Câmera: imagens com resolução de 1280 x 960 pixels, a 1Hz;

³¹ Isto é: estão referenciadas a um mesmo sistema de tempo global.

³² Peça de *software* responsável por integrar o equipamento ao ROS, por meio da leitura e interpretação de seus dados brutos, transmitidos por meio de uma interface de conexão (USB ou UART) e um protocolo de comunicação, geralmente serial. Além disso também lança uma janela que permite monitorar o estado da integração, avisando em uma eventual falha. Os drivers podem acessar arquivos *.xml* contendo configurações mutáveis para o acesso aos equipamentos, como configurações relativas a sistemas de referência e etc..

- b. IMU: orientação absoluta com relação ao norte magnético local; velocidade angular e aceleração livre no mesmo referencial da orientação ;
 - c. Receptor GNSS: posição (coordenadas geodésicas) e velocidade (ENU);
2. Ordenar a gravação dos dados em um arquivo *.bag*³³;
 3. Demandar a interrupção da gravação.

Todos esses comandos são executados por meio de scripts em linguagem *bash*³⁴, via acesso remoto, que por sua vez podem ser chamados por meio de atalhos clicáveis, de modo facilitar a interação do usuário. Todos os scripts e atalhos desenvolvidos, assim como a versão personalizada de arquivos de configuração estão disponíveis no repositório do Github mencionado em 3.1.

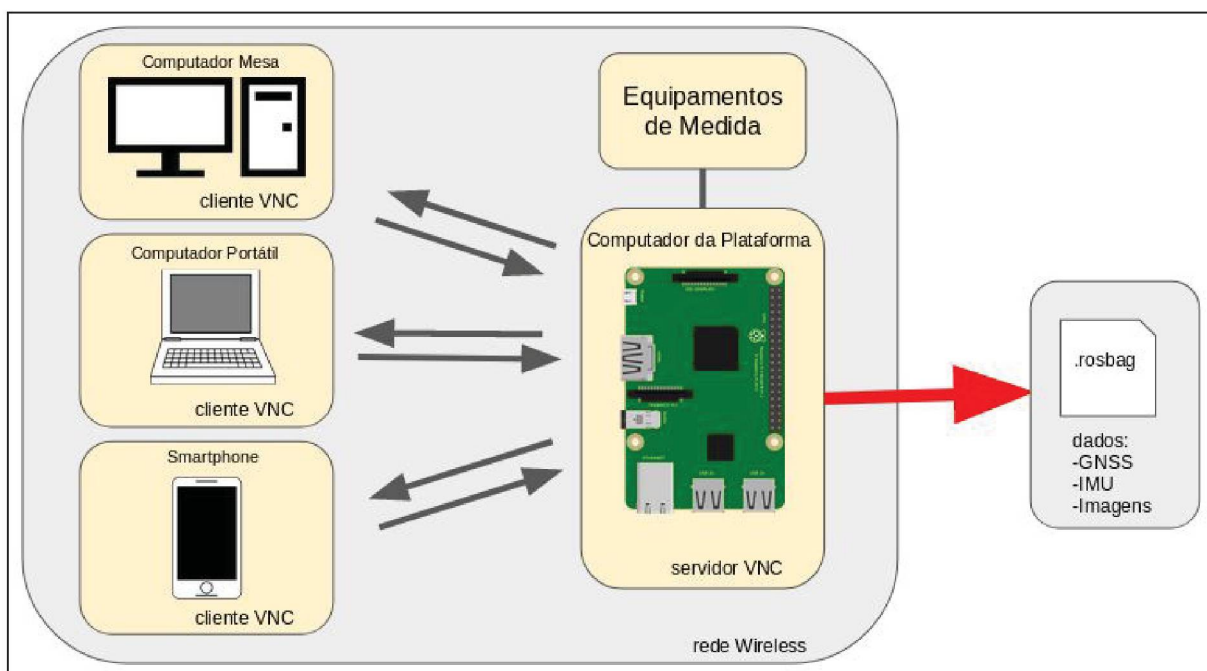
O acesso remoto é realizado por meio do protocolo VNC - *Virtual Network Computing*, que se utiliza do protocolo universal de “*buffer* remoto de telas”, que basicamente permite que um “cliente” controle a interface gráfica de um sistema operacional de um servidor, de modo que caso o cliente se desconecte, é possível que o mesmo se conecte novamente, tendo acesso ao mesmo estado no servidor (RICHARDSON & LEVINE, 2011).

Assim, o computador da plataforma mantém ativo um “servidor” VNC que pode ser acessado por meio de um “cliente” executado no dispositivo que o controlará. A conexão se dá por meio do compartilhamento de uma conexão de rede wireless, sendo que todos os dispositivos possuem adaptadores de rede com a capacidade para criação de rede sem fio. Assim, a plataforma pode ser controlada do interior do veículo sem a necessidade de cabos passantes pela janela do veículo. Os dispositivos podem ser: *desktop*, *notebook* ou *smartphone*. O esquema de conexão e coleta de dados estão demonstrados na figura 3.3.

³³ Formato próprio do ROS, responsável por armazenar dados advindos de diferentes sensores, em uma estrutura organizada que permite separar tais dados conforme necessário.

³⁴ Linguagem que permite criar *scripts* contendo uma sequência de comandos que seriam inseridos na linha de comando. Fonte: <https://en.wikipedia.org/wiki/Shell_script>.

Figura 3.3: Esquema de controle e gravação de dados do sistema apresentado



FONTE: O AUTOR (2018)

Assim, em toda e qualquer forma de processamento dos dados, será feito a leitura do arquivo “.rosbag” de saída, advindo da gravação. Com a execução de *scripts*, é possível realizar pré-processamento dos dados, ou extração dos mesmos em diferentes formatos. Para as aplicações do presente trabalho as imagens foram extraídas nativamente em formato “.jpeg”, e os demais dados em formato textual “.csv”.

3.2.2.2 Coleta de Dados

Todos os procedimentos descritos nessa seção consideram que a plataforma estará instalada em plataforma veicular terrestre do tipo “Carro”, como demonstrado na figura 3.4.

Figura 3.4: O SMMT Instalado em Plataforma Veicular Terrestre



FONTE: O AUTOR (2019)

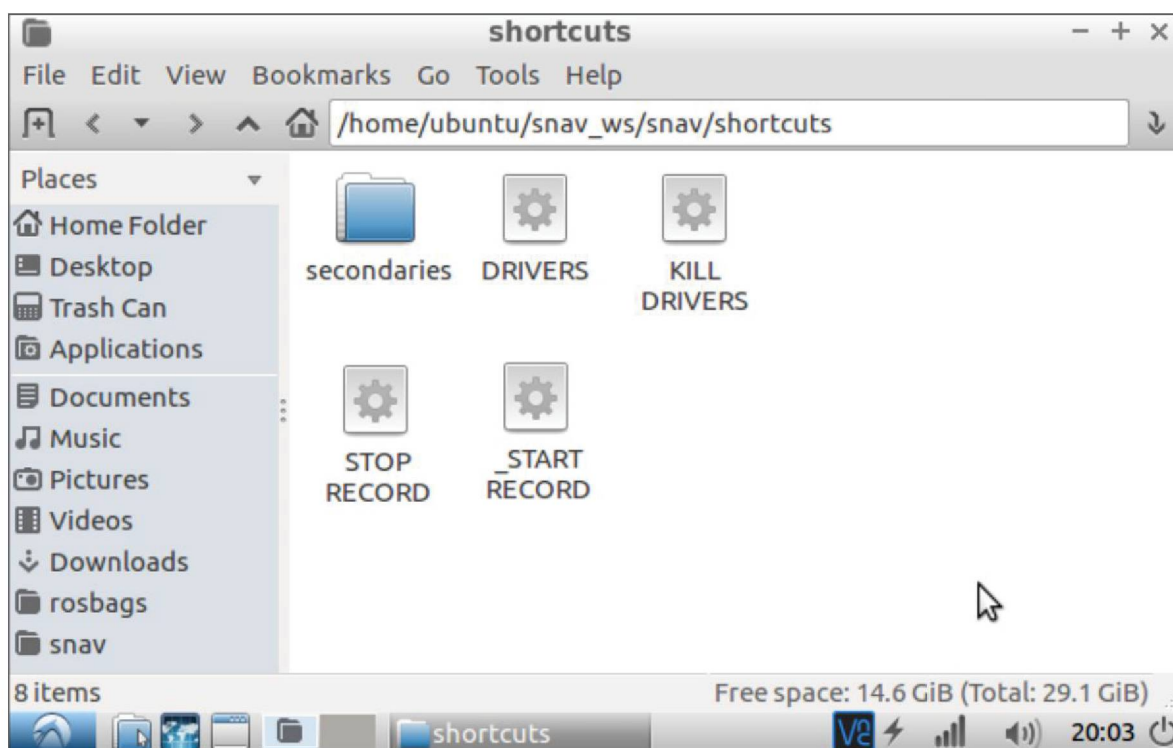
Inicialmente, deve-se planejar uma rota para a execução do levantamento. Para isto, diversos critérios devem ser levados em consideração, tais como (em ordem de prioridade):

- I. Área de interesse: deve-se realizar o levantamento na região da qual se deseja obter dados para realização do estudo;
- II. Cobertura: avaliação da validade da hipótese de que na execução da rota serão gravados todos os dados de interesse;
- III. Viabilidade: se a rota planejada não viola nenhuma regra de trânsito, como por exemplo, trafegar na contramão

Uma vez planejada a rota do levantamento, o sistema deverá ser instalado no topo do veículo. Após sua instalação, em local preferencialmente aberto (ainda que seja apenas para os padrões de um cenário altamente urbanizado), deverá ser acionado o interruptor que liga o computador da plataforma para inicialização da solução GNSS. Note que devem ser respeitados alguns minutos para que tal solução se estabilize³⁵.

Durante esse período é possível inicializar os *drivers*, como descrito na seção anterior. Em seguida é possível demandar o início da gravação dos dados. Ambos os *scripts* são chamados por meio de um *smartphone*, utilizado o protocolo VNC, tal qual mostrado na figura 3.4.

Figura 3.5: Interface utilizada para acionamento dos sensores da plataforma

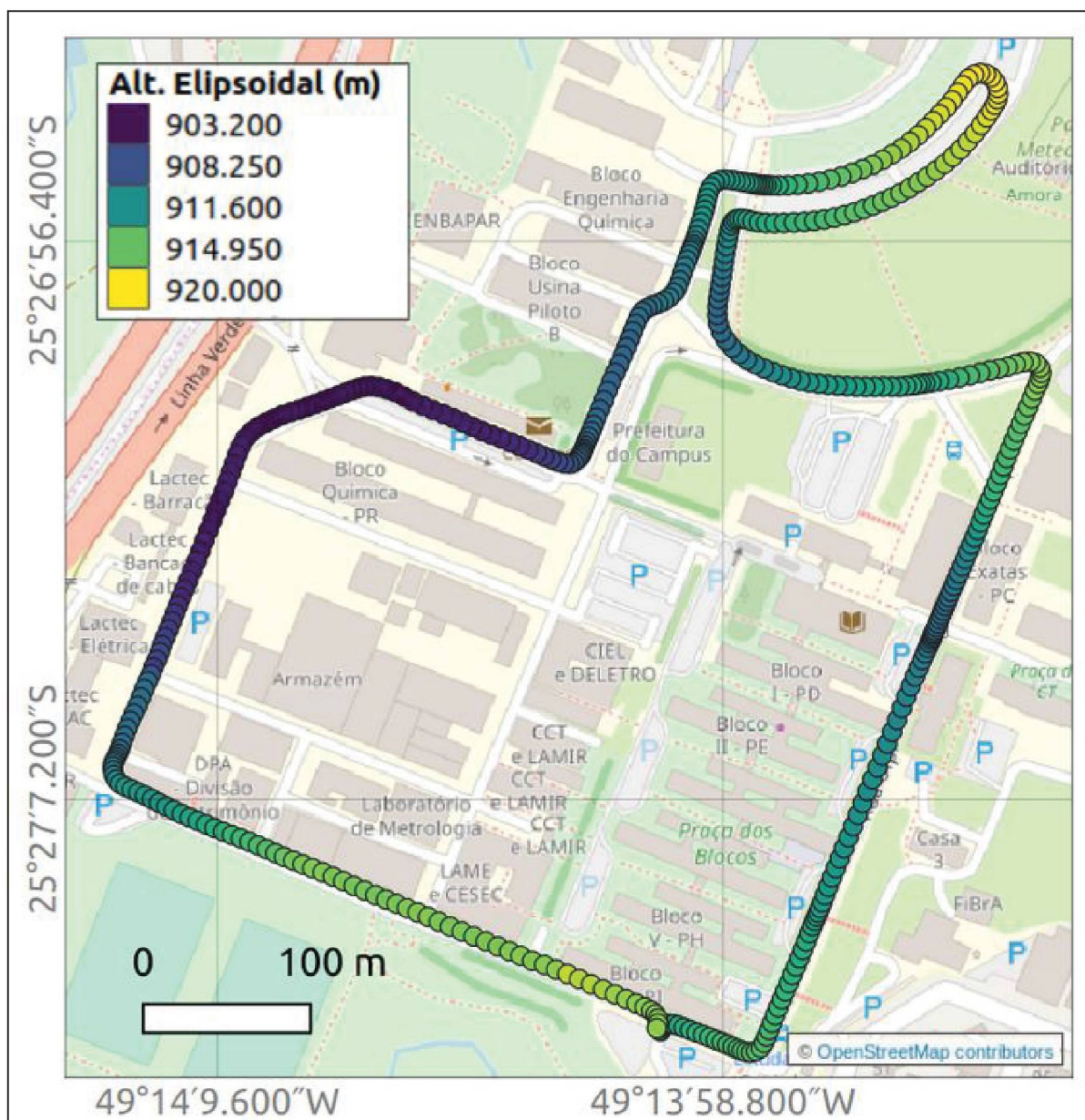


FONTE: O AUTOR (2019)

Após isso, a trajetória deverá ser executada. Na figura 3.6a é mostrado um exemplo de trajetória obtida.

³⁵ Conhecido como “*Time to First Fix*” (“Tempo até a primeira solução”), tempo necessário para que o receptor receba o mínimo de informações necessárias para calcular uma solução de posicionamento estável. Fonte: <<https://insidegnss.com/ready-to-navigate/>>.

Figura 3.6a : Trajetória percorrida pelo veículo em um teste de exemplo



FONTE: O AUTOR (2019); Mapa-Base: OpenStreetMap

Uma vez executado o levantamento, os dados poderão ser descarregados por meio de um computador de mesa, seja fazendo *upload* dos dados em um serviço de armazenamento na nuvem ou copiando-os para um dispositivo de armazenamento ou para um dispositivo disponível na mesma rede local. Como já descrito anteriormente, todos os dados são gravados em um único arquivo ".rosbag".

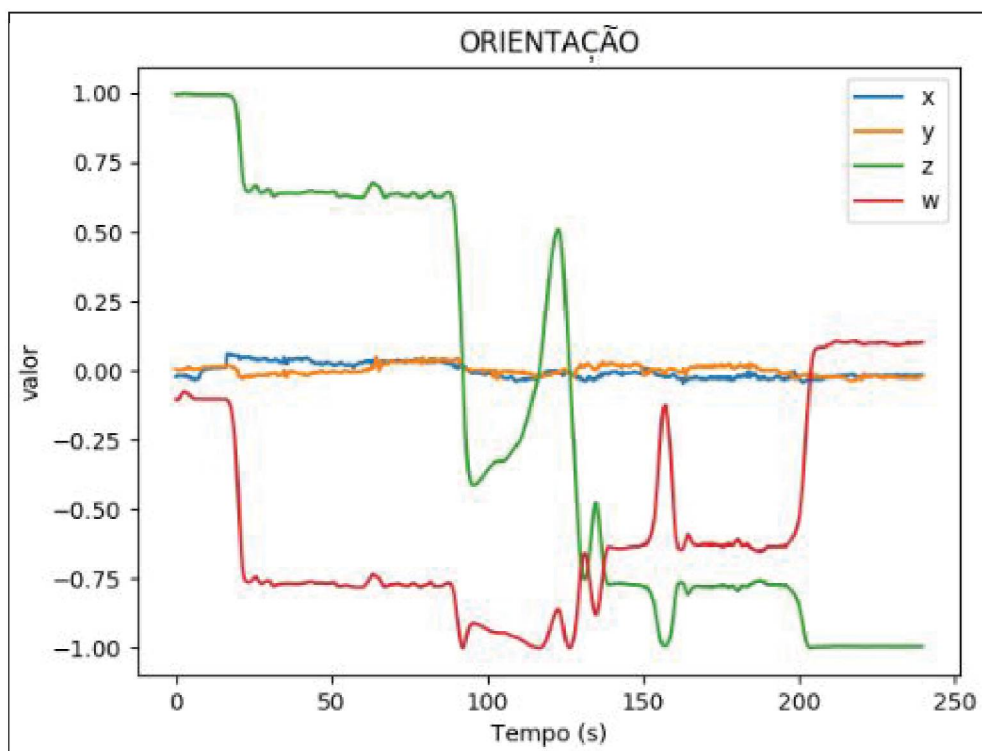
3.2.3 Processamento dos dados

O presente subitem será dividido em duas seções:

1. Treinamento e uso das redes neurais convolucionais: Serão descritos os procedimentos para obtenção dos modelos para segmentação de vegetação x não-vegetação³⁶, assim como sua aplicação em imagens.
2. Processamento dos dados obtidos pela plataforma em levantamentos: serão apresentados os procedimentos detalhados para os tópicos introduzidos em 3.1.

Vale salientar que, uma série de *scripts* para a extração dos dados é executada. Como o arquivo *.rosvbag* é um formato binário, os dados devem ser primeiramente decodificados. Na figura 3.6b é apresentado um exemplo de gráfico que representa os dados extraídos da IMU. O leitor poderá notar que, os dados estão exibidos na parametrização nativa para a orientação, isto é, em quatérnios unitários.

Figura 3.6b: gráficos com dados extraídos da IMU

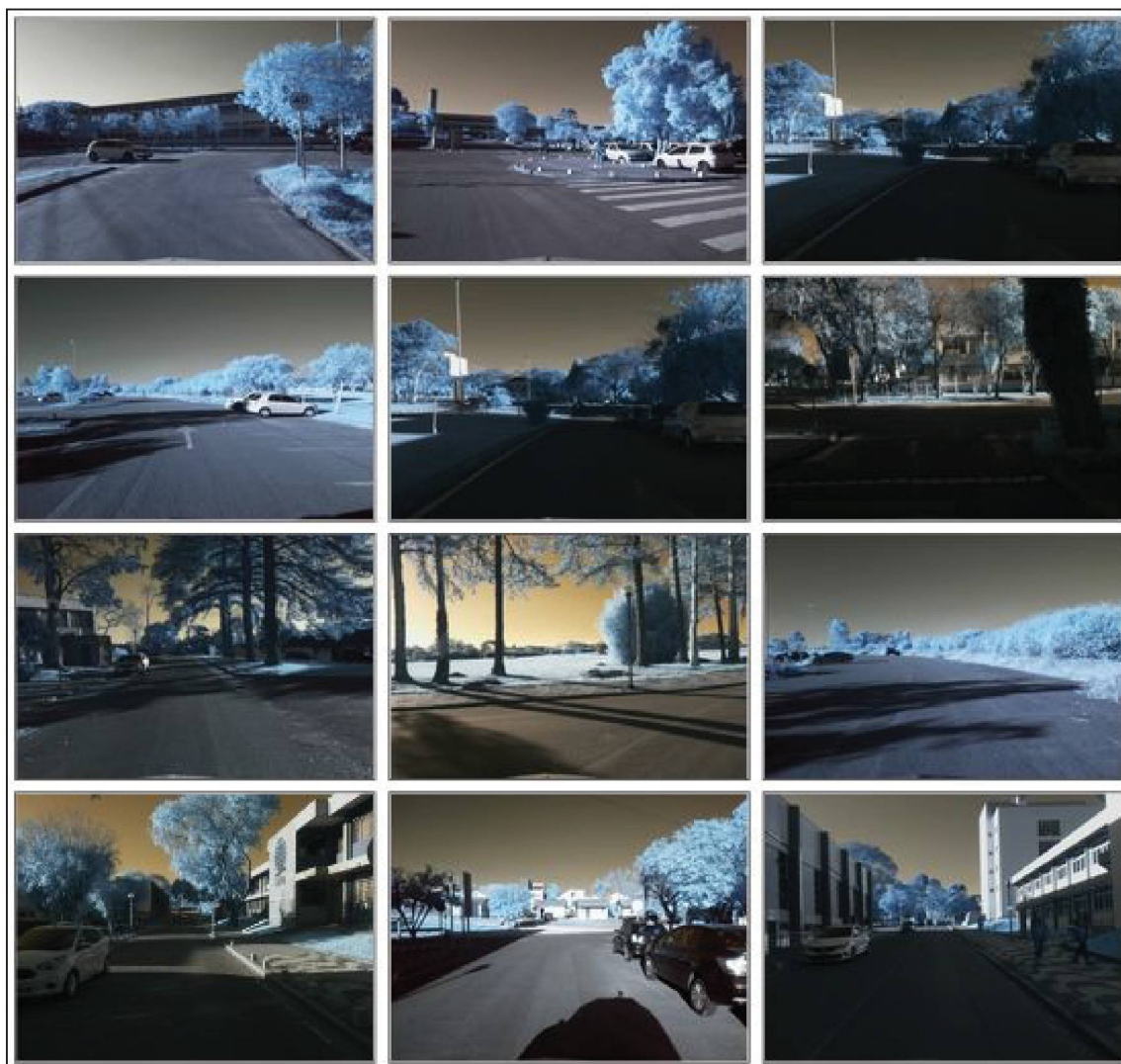


³⁶ Alguns modelos foram treinados para a classificação apenas da vegetação não pertencente ao terreno, e outros para qualquer tipo de folhagem que tenha sido observado no conjunto de treinamento..

FONTE: O AUTOR (2019)

Na figura 3.7 são apresentados alguns exemplos de imagens tomadas pela câmera embarcado na plataforma terrestre.

Figura 3.7: Amostra de imagens tomadas pela plataforma

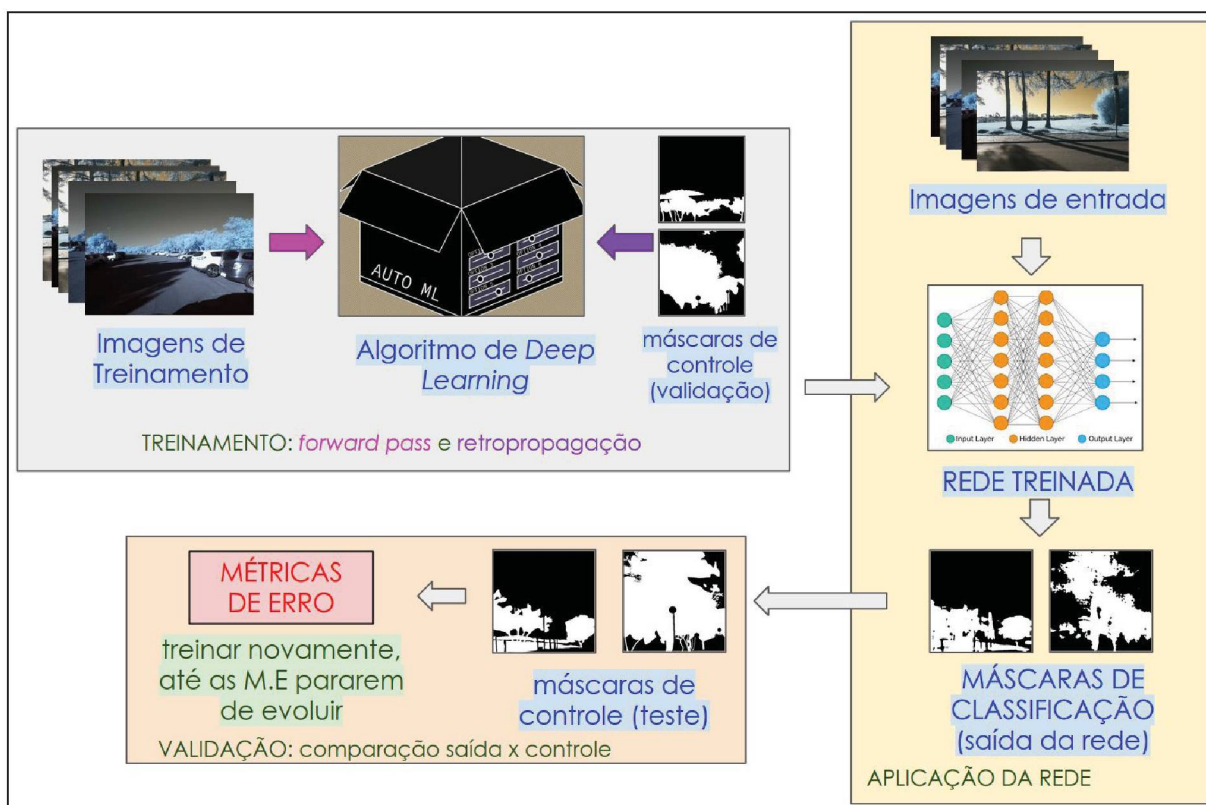


FONTE: O AUTOR (2019)

3.2.3.1 Treinamento e Uso das Redes Neurais Convolucionais

Um esquema das CNNs empregadas neste trabalho, está apresentado na figura 3.8.

Figura 3.8: esquematização do uso de CNNs no presente trabalho



FONTE: O AUTOR (2019);

Os dados amostrais para o treinamento da rede consistem em:

1. Pares de imagens com (a) Fotografia tomada em cenários de interesse (no caso do presente trabalho, consistem de cenas urbanas tomadas nas vias públicas); e (b) Classificação das categorias presentes na fotografia, também chamadas de “máscaras”, sendo atribuída a cada classe uma cor única. Neste trabalho, tal classificação é realizada manualmente.
2. Um dicionário contendo os nomes das classes, associando-os às suas respectivas cores nas máscaras manualmente classificadas.

Os pares de imagens são divididos em três grupos distintos, como segue:

1. *Train* (“Treinamento”): imagens utilizadas diretamente pelo algoritmo de *Deep Learning* para modelagem dos parâmetros da Rede Neural;
2. *Validation* (“Validação”): imagens utilizadas pelo algoritmo de *Deep Learning* para verificação interna da qualidade e reajuste de seus parâmetros;

3. *Test* (“Teste”): imagens não empregadas pelo algoritmo de *Deep Learning*, utilizadas como controle externo para verificação da qualidade do modelo obtido;

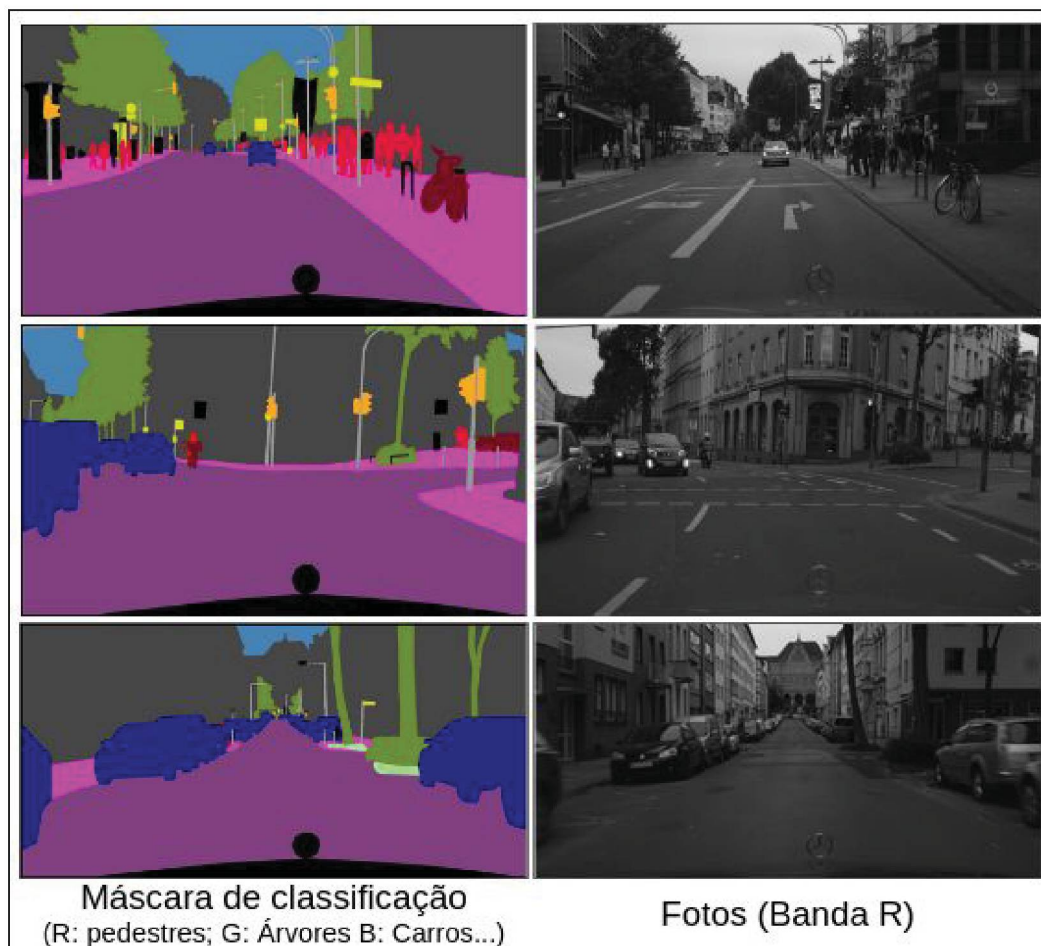
Tais dados são distribuídos em conjuntos de dados (“*Datasets*”), sendo possível gerá-los com dados próprios ou obtê-los por meio de *download* em bases de dados públicas (“*benchmarks*”). Para o desenvolvimento do presente trabalho foram testados dois conjuntos diferentes, cada qual com suas particularidades listadas:

1. O *Cityscapes Dataset* (CORDTS, et. al., 2016): contém cenas urbanas de mais de 50 cidades européias, com milhares de pares de imagens e máscaras de classificação:
 - Não foram criados objetos com “buracos”, por exemplo, se uma árvore está na frente de uma fachada, as partes visíveis de tal fachada não foram classificadas como “muro” (na região onde há folhas);
 - A vegetação pertencente ao terreno é classificada na categoria “terreno”
 - Cerca de 30 classes, envolvendo: carros, vegetação, postes, placas, pessoas etc.
2. Conjunto com 50 imagens tomadas pelo SMMT, classificadas manualmente:
 - Apenas duas classes: vegetação e não vegetação
 - Mesmo objetos que estão no fundo da vegetação, estão classificados como “não-vegetação”, isto é, foram criados objetos com buracos.
 - Foi compilada em duas versões: uma contendo vegetação do terreno e outra não contendo-a.

Na figura 3.9 é apresentado um exemplo das amostras de pares de imagens e máscaras do Dataset Cityscapes. O *Dataset* foi originalmente gerado com uma câmera capaz de capturar todo o espectro na região do visível. Entretanto, uma vez

que a informação das bandas G e B não estão disponíveis nas imagens tomadas pelo SMMT. Assim sendo, apenas a banda R do *dataset Cityscapes* foi utilizada para o treinamento dos cenários de treinamento em que se utilizou tal *dataset* para tal finalidade.

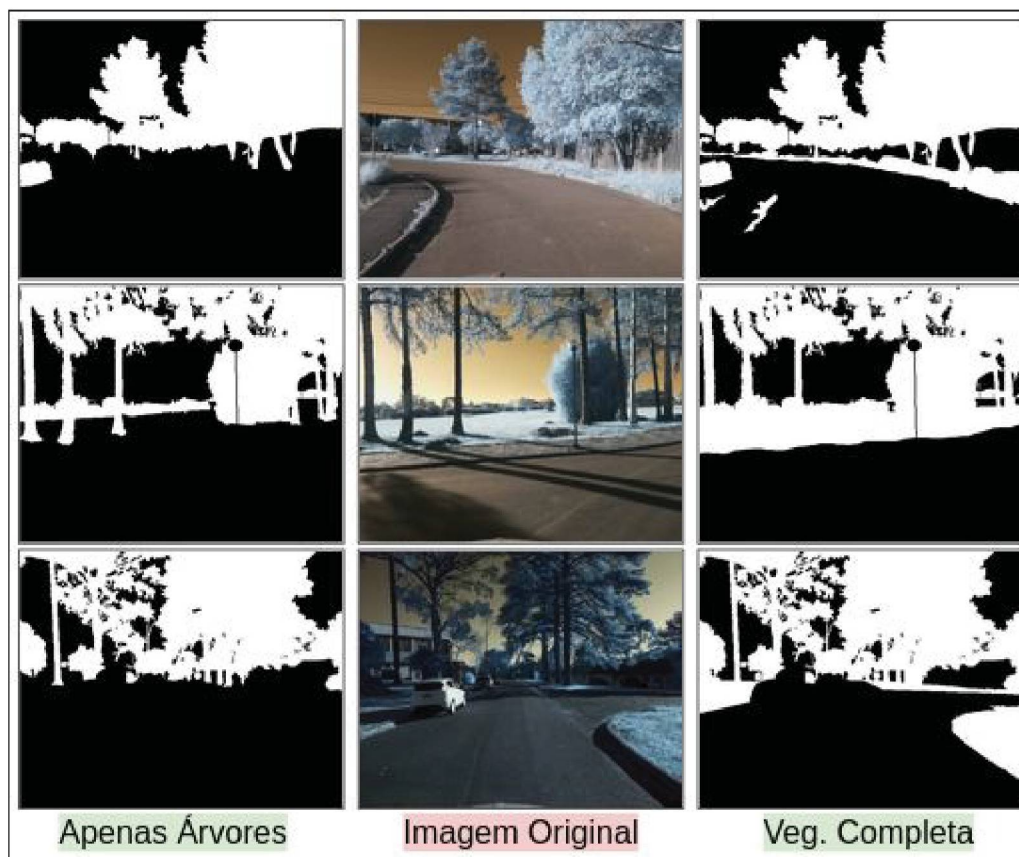
Figura 3.9: Pares de Imagens disponibilizados no *Cityscapes Dataset*



FONTE: CORDTS, et. al. (2016)

Na figura 3.10 estão alguns exemplos de máscaras de classificação de vegetação e não-vegetação, nas duas categorias utilizadas.

Figura 3.10: Amostras de máscaras de classificação de imagens tomadas pela plataforma



FONTE: O AUTOR (2019)

Uma vez disponíveis os dados, os mesmos foram separados na seguinte proporção: 60% treinamento; 20% Validação; e 20% teste. Após a separação dos dados, é empregado o *framework* denominado *Semantic Segmentation Suite* (SEIFF, 2017).

É possível escolher uma série de opções para treinamento de uma rede neural a partir de um conjunto de dados adequado ao formato de entrada. Tais opções³⁷ se dividem em dois grandes grupos:

1. Pré-Processamento das imagens:

- a. Redimensionamento das imagens: reduz o tempo de processamento da rede como um todo, entretanto, também reduz a qualidade da segmentação, conforme é piorado a resolução da imagem. Também pode ser utilizado para adequar

³⁷ As opções aqui mencionadas são as presentes no conjunto de ferramentas escolhidas para a realização do presente trabalho, no entanto não há limitações para operações de pré-processamento de imagens. Assim como é possível personalizar mais uma rede neural convolutiva.

a imagem às dimensões dos filtros convolutivos a serem utilizados, assim, geralmente as imagens são transformados para potências de 2. Neste trabalho, as imagens foram reduzidas de 1280x960 para 512x512, devido às limitações dos recursos computacionais disponíveis.

- b. Operações geométricas/radiométricas nas imagens: operações de espelhamento, rotação, mudança de brilho são utilizadas para “Ampliação dos Dados” (*Data Augmentation*), que consiste em realizar transformações (com variações máximas pré-estabelecidas) de modo a possibilitar que a rede observe as propriedades geométricas e cromáticas³⁸, dos objetos com menor influência da posição do sensor e de como o mesmo capta a energia eletromagnética (MIKOŁAJCZYK & GROCHOWSKI, 2018). No presente trabalho, utilizou-se apenas uma variação aleatória de até 10% no brilho.

2. Customização da rede neural escolhida (hiperparâmetros):

- a. Número de épocas de treinamento: uma época é um ciclo completo de processamento de todas as imagens disponíveis. A qualidade da segmentação evolui ao longo das épocas, eventualmente parando num determinado patamar. Nos modelos testados, neste trabalho, não se introduziu objetivamente um limite para o número de épocas.
- b. Tamanho do “lote” de dados ("*batches*"): geralmente o conjunto completo de dados é muito denso, gerando custos computacionais acima dos recursos requeridos. Desta forma, o processamento é feito com pequenos lotes de dados. Para os testes executados na máquina local, foram utilizados lotes com tamanho de 1 a 3 imagens e para os executados nas máquinas remotas do *Google Colab*, de 10 a 40 imagens.

³⁸ Estas que de fato, são local e instantaneamente invariáveis para o objeto de estudo, na realidade tridimensional.

- c. *Frontend*: modo que a rede age dentro de cada camada ou como as interliga. Para maiores detalhes, vide a seção 2.6. No presente trabalho foram utilizados os modelos: *ResNet50*; e *InceptionV4*.
- d. Modelo: a estrutura de camadas da rede, infinitas combinações são possíveis. Para os tipos de camadas, vide a seção 2.6. Vários modelos foram testados.

A cada época de treinamento completada, é possível salvar o estado da rede em um “ponto de checagem” (*checkpoint*), a partir do qual já é possível processar uma imagem, produzindo a segmentação de saída.

O desempenho da rede é mensurado por meio de métricas de erro obtidas à partir de diferentes comparações de uma imagem gerada como saída da rede (classificação predita) com aquela classificada manualmente (classificação de controle).

Figura 3.11: Possibilidades para a comparação de um pixel predito e da amostra de controle

		CONTROLE	
		VEGETAÇÃO (1 ou VERDADEIRO)	NÃO VEGETAÇÃO (0 ou FALSO)
PREDIÇÃO	VEGETAÇÃO (1 ou VERDADEIRO)	VERDADEIRO POSITIVO (VP)	FALSO POSITIVO (FP)
	NÃO VEGETAÇÃO (0 ou FALSO)	FALSO NEGATIVO (FN)	VERDADEIRO NEGATIVO (VN)

FONTE: O AUTOR (2019)

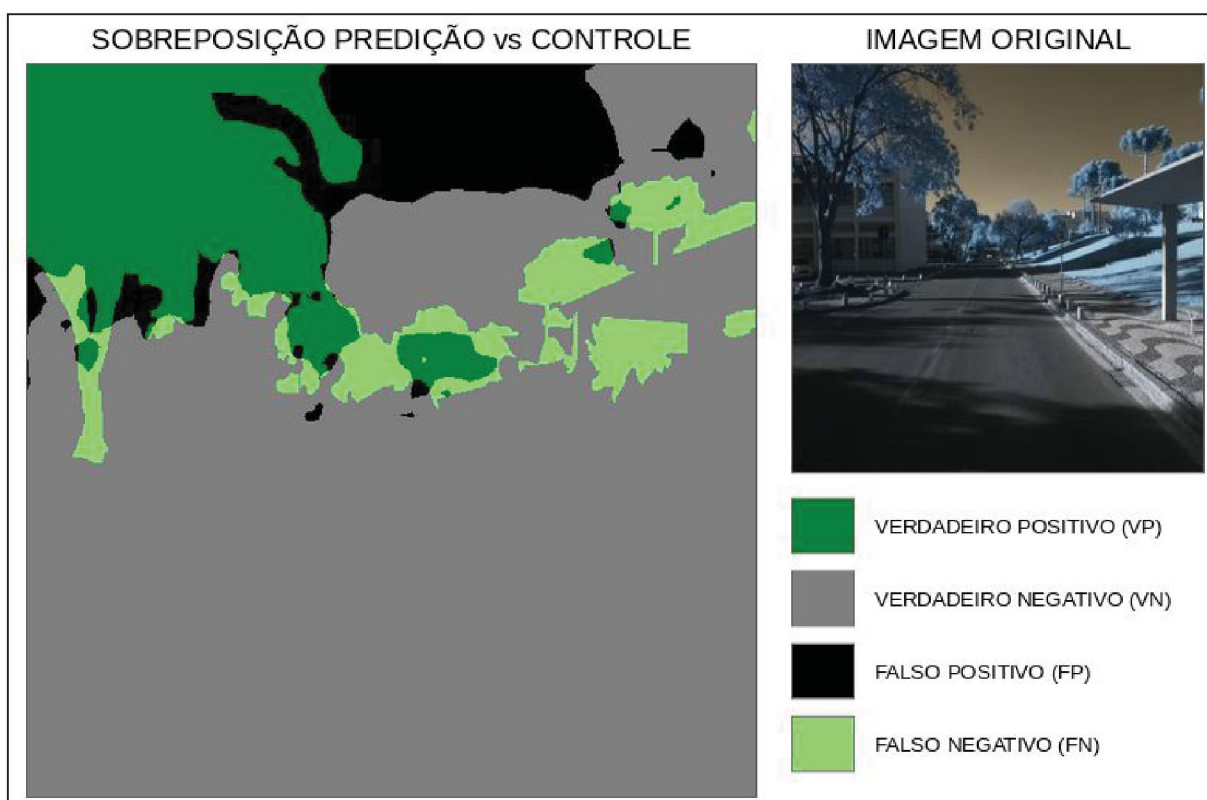
Antes de mencioná-las, é necessário explicar as possibilidades envolvendo a comparação pixel-a-pixel das duas imagens, isto é, o pixel (0,0) da imagem predita será comparado com o pixel de mesmo índice da imagem de controle e assim por diante até a comparação entre os últimos pixels das duas imagens. Na imagem 3.11

estão apresentadas as quatro possibilidades de comparação para o objetivo do presente trabalho (para a classificação de vegetação e não vegetação).

A figura 3.11 apresentada acima é muito semelhante à chamada “Matriz de Confusão” (FAWCETT, 2006). Vale lembrar que considerando o objeto de estudo deste trabalho, um Falso Positivo é considerado um erro mais deletério que um Falso Negativo³⁹, uma vez que objetos que não pertencem a vegetação poderão ser alvo de análises automáticas posteriores, as quais deveriam ser executadas apenas sobre a vegetação.

Um exemplo com as quatro possibilidades, visualmente demonstrada por meio da sobreposição de uma predição⁴⁰ com uma amostra de controle é exibido na figura 3.12.

Figura 3.12: Exemplo ilustrativo de predição sobreposta a imagem de controle



FONTE: O AUTOR (2019)

³⁹ Neste caso, a negligência do fenômeno estudado torna-se menos danosa, dada a extensão das malhas viárias urbanas.

⁴⁰ Foi escolhida propositalmente uma imagem resultante do processamento utilizando como modelo os pesos de uma época do início (com desempenho inquestionavelmente limitado) de um treinamento, para uma ilustração mais elucidativa.

Uma vez que o modelo (a rede) realiza a predição de vegetação e não vegetação ou uma predição contendo várias classes a ser transformada em apenas vegetação e não vegetação, é realizado o somatório de pixels para cada uma das quatro possibilidades (VP, VN, FP e FN). A partir dos somatórios, são calculadas as métricas de erro. Há uma grande variedade de métricas de erro passíveis de serem utilizadas, entretanto, para as análises aqui realizadas foram escolhidas cinco delas:

- Taxa de Acerto (TA)⁴¹: apresenta a razão entre os acertos e o total (FAWCETT, 2006), conforme apresentado na equação 3.1.

$$TA = \frac{\text{acertos}}{\text{total de pixels}} = \frac{VP + VN}{VP + FP + VN + FN} \quad (3.1)$$

- Valor Preditivo Positivo (VPP)⁴²: apresenta a razão entre os acertos relevantes e as predições realizadas (FAWCETT, 2006), conforme apresentado na equação 3.2.

$$VPP = \frac{VP}{\text{preditos como vegetação}} = \frac{VP}{VP + FP} \quad (3.2)$$

- Sensibilidade (S): apresenta a razão entre os acertos relevantes e os pixels que realmente são vegetação (adaptado de FAWCETT, 2006), conforme apresentado na equação 3.3.

$$S = \frac{VP}{\text{realmente vegetação}} = \frac{VP}{VP + FN} \quad (3.3)$$

⁴¹ Na literatura ligada à área de *Machine Learning* é denominada de *accuracy*, entretanto, considerando o âmbito das Ciências Geodésicas, a tradução direta, “acurácia”, possui um significado não totalmente compatível com o dessa métrica, assim foi optado por uma tradução livre.

⁴² Muito semelhante ao caso anterior, originalmente chamada “*precision*”, foi buscada uma nomenclatura semelhante, traduzida livremente, devido às mesmas questões do caso anterior.

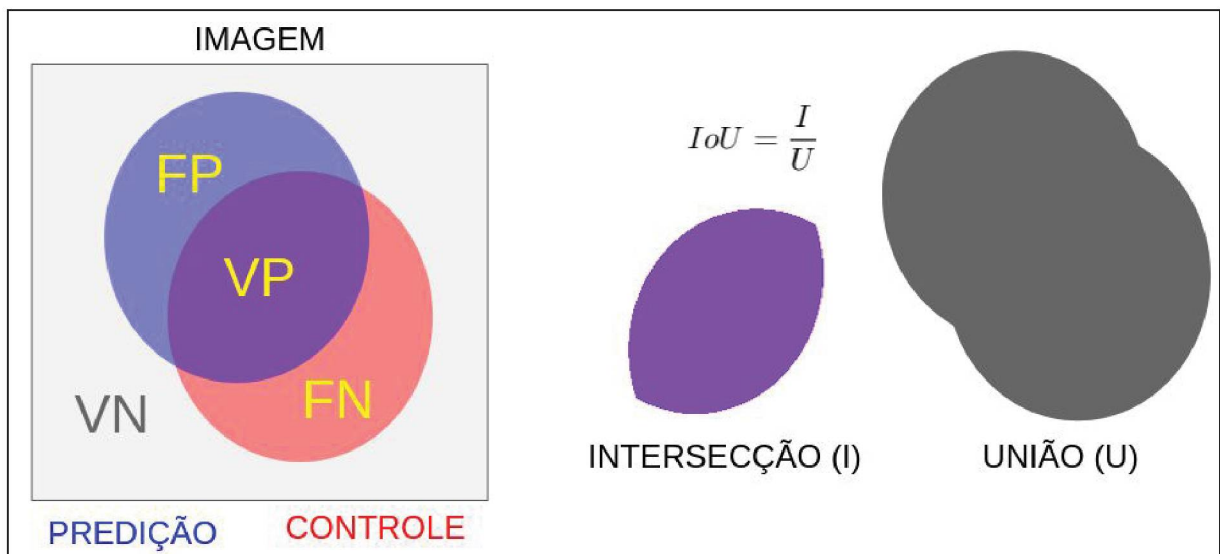
- Score F1: é a média harmônica entre a Sensibilidade e o VPP (FAWCETT, 2006), conforme apresentado na equação 3.4.

$$F1 = 2 * \frac{VPP * S}{VPP + S} = \frac{2VP}{2VP + FP + FN} \quad (3.4)$$

- Razão Intersecção/União (IoU): é a razão dos pixels corretamente classificados como vegetação por todos os pixels relevantes para o problema (RAHMAN & YANG, 2016), como apresentado na equação 3.5 e demonstrado de maneira conceitual na figura 3.13.

$$IoU = \frac{\text{intersecção}}{\text{união}} = \frac{VP}{VP + FP + FN} \quad (3.5)$$

Figura 3.13: demonstração conceitual da métrica “IoU”

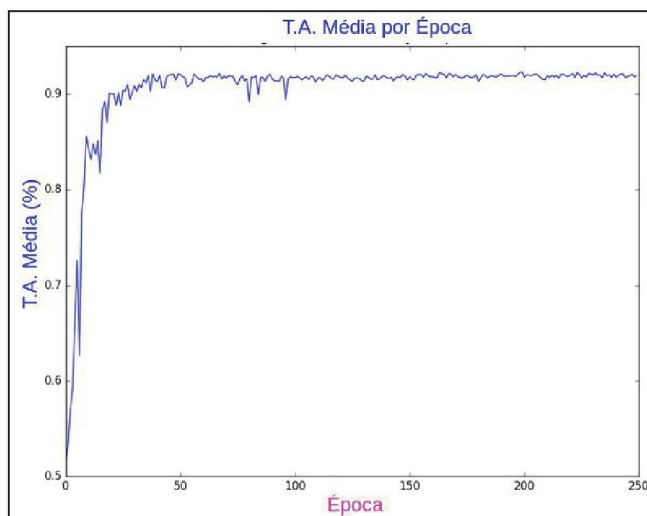


FONTE: O AUTOR (2019)

Ao longo das épocas é possível acompanhar a evolução das métricas de erro. As mesmas apresentam um crescimento, sobretudo, após as primeiras épocas. Geralmente as evoluções possuem um comportamento que remete a uma função de

logaritmo natural, um exemplo de evolução de métrica de erro (TA) está apresentado na figura 3.14.

Figura 3.14: Exemplo da evolução da Taxa de Acerto (TA)



FONTE: Adaptado de SEIFF(2017)

Como pode ser observado na figura 3.14, após uma determinada época, a esperança matemática da taxa de variação da TA se mantém estável. Assim, após a verificação de tal fato é possível parar a execução do algoritmo de *Deep Learning* para treinamento da rede neural.

Uma vez encerrado o treinamento, para cada época ficam armazenados os modelos (cada qual composto por pesos e tendências) que podem ser acoplados à CNN para execução da classificação, resultando no desempenho referente àquela época. De posse da lista de modelos, é possível escolher o melhor a partir das métricas de erro. No presente trabalho se utilizou empiricamente como melhor modelo aquele que apresentou um maior número de máximo⁴³ desempenho em um número maior de métricas.

3.2.3.2 Processamento dos Dados Obtidos pelo SMMT

Após a obtenção dos dados de saída, como já explicado em 3.2.1, é executada a extração dos dados gravados em um arquivo ".rosbag", por meio de

⁴³ Isto é, melhor que qualquer outro modelo.

scripts em linguagem *bash*, a partir do qual os dados são extraídos da seguinte maneira:

- I. Imagens (frequência de captura variada), extraídas e nomeadas de maneira ordinal em formato “.jpeg” e posteriormente transformadas para o formato “.png”⁴⁴;
- II. Dados GNSS (1 Hz), extraídos para um arquivo “.csv”, contendo: posição (latitude, longitude, altitude elipsoidal); status da navegação; MVC; tipo da MVC (aproximação baseada nos DOPs, ou obtida rigorosamente);
- III. Dados da IMU (60 Hz), extraídos para um arquivo “.csv”, contendo os seguintes dados: aceleração linear livre (m/s^2); velocidade angular (rad/s); orientação (quatérnio unitário);

Todos os dados extraídos estão acompanhados da anotação do instante de tempo em que foram capturados. Tal sistema de tempo é materializado pelo ROS e se trata do “tempo UNIX”, cujo segundo 0 é à meia noite do dia 1 de janeiro de 1970 (MILLS 2002).

Os dados GNSS e da IMU não são capturados de maneira sincronizada com as imagens da câmera. Desta forma, considerando o deslocamento da plataforma, as observações não serão capturadas com o veículo na mesma posição do instante da tomada da fotografia⁴⁵. Assim, se faz necessário um processo de interpolação, para cada caso foram feitas as seguintes considerações:

- A. IMU: uma vez que orientação é dada na forma de um quatérnio de módulo 1 e qualquer alteração nos valores pode violar tal característica⁴⁶, produzindo um valor que não corresponde à rotação desejada. Posto isso, considerando-se a alta frequência de captura dos dados da

⁴⁴ O Formato “jpeg” apresenta artefatos advindos da compressão, não preservando os *bitmaps* completos, assim para o treinamento da rede, se faz necessária a conversão para o formato “png”.

⁴⁵ Lembrando que há deslocamentos espaciais e desalinhamentos entre os sistemas referenciais dos diferentes equipamentos.

⁴⁶ Seria necessária uma interpolação realizada na hiper-esfera quadridimensional de raio 1.

IMU, considerou-se que o uso do “vizinho-mais-próximo” seria o suficiente, uma vez que o maior intervalo de tempo possível entre a captura de uma imagem e de um conjunto da IMU será de 0,05 s.

B. GNSS: como a posição é tratada como um vetor em um espaço euclidiano, é possível aplicar uma interpolação linear sobre suas componentes, contanto que estejam de fato expressas como um vetor. Assim, após transformar a posição em coordenadas geodésicas para um vetor no ECEF ou para um SGL, é possível aplicar, sobre suas componentes X,Y e Z a interpolação linear, cuja fórmula é apresentada na equação 3.6.

$$f(t_1 + \Delta t_1) = f(t_1) + \frac{[t_1 + \Delta t_1] - t_1}{t_2 - t_1} [f(t_2) - f(t_1)] \quad (3.6)$$

Onde:

- $f(t_1 + \Delta t_1)$ é o valor que se deseja interpolar, equivalente a $f(t_2 - \Delta t_2)$;
- $f(t_1)$ é a observação em momento anterior à tomada da fotografia;
- $f(t_2)$ é a observação em momento posterior à tomada da fotografia;
- t_1 e t_2 são os tempos em que as observações anteriores e respectivamente posteriores, ao instante da tomada da fotografia, foram registradas; e
- $[t_1 + \Delta t_1]$ é o instante de tempo em que a fotografia foi tomada

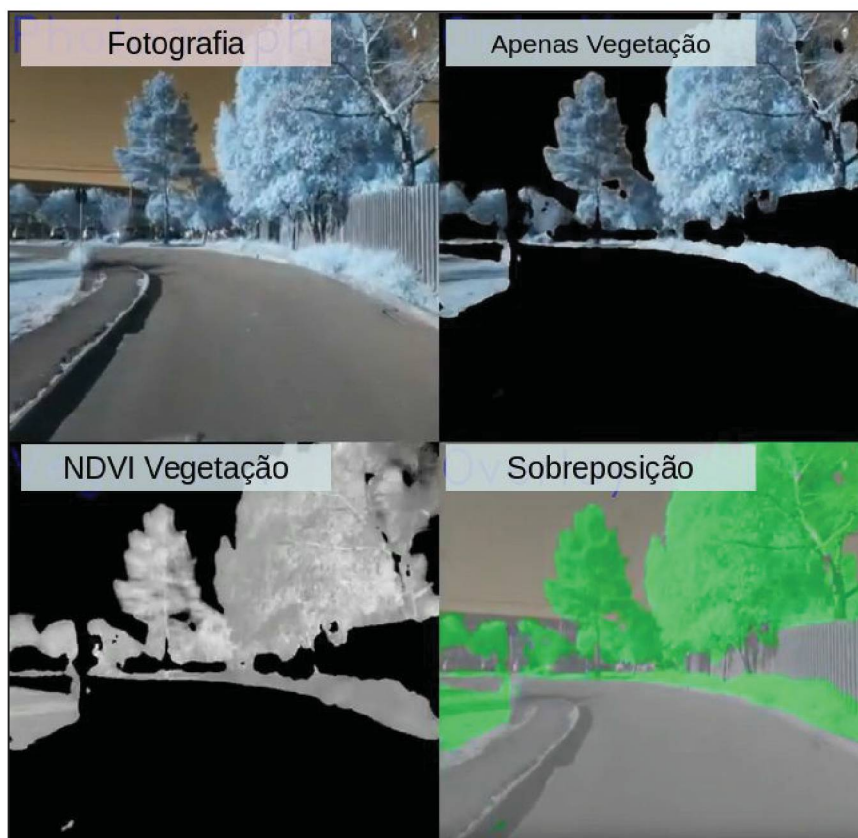
Uma vez que o treinamento da rede neural tenha sido executado, e a melhor época (que estará entre as últimas) escolhida por meio do processo de votação já mencionado, é possível passar um conjunto inteiro de imagens pela rede neural, para isso foram gerados *scripts* em linguagem *bash*.

Como já introduzido em 3.2.3.1, a saída da rede neural é uma imagem chamada “máscara de classificação”, onde cada cor corresponde a uma classe diferente. No caso de uma classificação apenas vegetação e não-vegetação trata-se de uma imagem binária onde os pixels brancos são os classificados como vegetação

e os pixels pretos como sendo de qualquer outra classe. A partir da imagem original e da máscara⁴⁷, é possível gerar uma série de imagens:

1. Recorte apenas vegetação: gerada por meio da aplicação de um operador booleano na imagem, usando a máscara de classificação como entrada;
2. Imagem de NDVI da vegetação: sobre a imagem obtida no item anterior será aplicada sobre seus pixels válidos a equação 2.1;
3. Imagem de sobreposição: gerada por meio da combinação linear da máscara de classificação (com as cores substituídas por cores mais representativas) e da imagem original, é útil para inspeção visual do resultado obtido;

Figura 3.15: Fotografia e imagens geradas a partir dela e da classificação



FONTE: O AUTOR (2019)

⁴⁷ Sempre binarizada, caso a rede tenha sido treinada como o *dataset Cityscapes*, todas as demais cores serão substituídas por preto e a classe vegetação por branco.

Um exemplo dos tipos de imagens mencionadas, juntamente com uma imagem original estão apresentadas na figura 3.15.

A imagem contendo apenas a vegetação é propícia para a realização de qualquer análise ligada a vegetação, sendo que há uma série de índices de vegetação que podem ser utilizados para análises. No presente trabalho, optou-se pelo emprego do NDVI, por ser o índice mais amplamente utilizado para estudos de vegetação, sobretudo a respeito da folhagem (JIANG et. al. 2006).

4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

No presente capítulo serão apresentados os experimentos e a análise dos resultados obtidos com os procedimentos metodológicos aplicados sobre os dados coletados.

4.1 EXPERIMENTOS REALIZADOS

Para a avaliação da metodologia proposta, foram realizados 2 experimentos:

- I. Coleta completa de dados com o SMMT;
- II. Treinamento de CNNs para posterior validação, de modo a ser possível determinar qual a rede com melhor desempenho;

Para o experimento I, o local escolhido foi o Campus Centro Politécnico da UFPR. Os parâmetros de captura utilizados estão apresentados na tabela 4.1.

Tabela 4.1: Parâmetros de captura dos dados do SMMT utilizados no experimento

Equipamento	IMU	Ant/Recep. GNSS	Câmera
Frequência (Hz)	100	1	10
Constelações	N/A	GPS/GALILEO/GLONASS	N/A
Modo de Captura	N/A	Automotivo ⁴⁸	N/A
Resolução	N/A	N/A	1280 x 960
Tempo de Exposição (s)	N/A	N/A	1/1000

FONTE: O AUTOR (2020)

A rota executada corresponde à rota apresentada na figura 3.6a, foram 2,1 Km percorridos, ao longo de 11,9 minutos com 7111 fotografias capturadas, em um conjunto de dados com cerca de 4,6 GB.

⁴⁸ Aplica restrições ao cômputo da solução, já que o carro é uma plataforma não-holonômica.

Com relação ao experimento II, que se trata de fato de um conjunto de experimentos, foram testados vários cenários para o treinamento das redes, cujos parâmetros gerais estão apresentados na tabela 4.2.

Tabela 4.2: resumo dos cenários analisados para redes neurais

	Cenário A		Cenário B
	I	II	
Classificação	Apenas Árvores		Toda a Vegetação
Avaliação	Toda a Vegetação	Apenas Árvores	Toda a Vegetação
Dataset Tr.	Cityscapes		Próprio
Dataset Val.	Cityscapes		Próprio
Dataset Teste	Próprio	Próprio (apenas árvores)	Próprio
Resolução	512 x 512		
V. Brilho Aleat.	10%		
Número de Épocas	aprox 200*		aprox. 1000
Tamanho do Lote	1		20
Imagens Tr./Val..	3000		40
Imagens Teste	50		10
Frontend	ResNet50		InceptionV4
Modelo de Rede	DeepLabV3+		FRRN

FONTE: O AUTOR (2020)

Para os *datasets* próprios, foram escolhidas aleatoriamente 50 imagens, advindas do experimento I. Houve grandes diferenças em termos de *hardware*: para o cenário A, utilizou-se um computador limitado em recursos, em que cada época levou cerca de 12 horas para ser treinada (cerca de 3 meses de treinamento); ao passo que para o cenário B, foi empregada uma máquina remotamente disponibilizada pelo sistema *Google Colab*, onde cada época levou cerca de 5 minutos para ser treinada, isto é, cerca de 4 dias de treinamento.

4.2 RESULTADOS E ANÁLISES

O primeiro resultado a ser abordado é a plataforma em si, sendo que parte do presente trabalho foi a escolha de seus componentes e montagem. Para a adequada captura dos dados, a estabilidade atingida pela plataforma se mostrou fundamental. Ainda que não se tenha elaborado uma maneira objetiva de avaliar o êxito da plataforma nas tarefas de estabilidade, capacidade de controle, gravação e transmissão de dados, evidentemente, um eventual insucesso em tais tarefas, teria de fato inviabilizado as etapas futuras.

4.2.1 ANÁLISE DOS DADOS DE NAVEGAÇÃO

No tocante aos dados de navegação, uma vez que não foi implementada a fusão de sensores, serão feitos comentários relativos à orientação e posição separadamente.

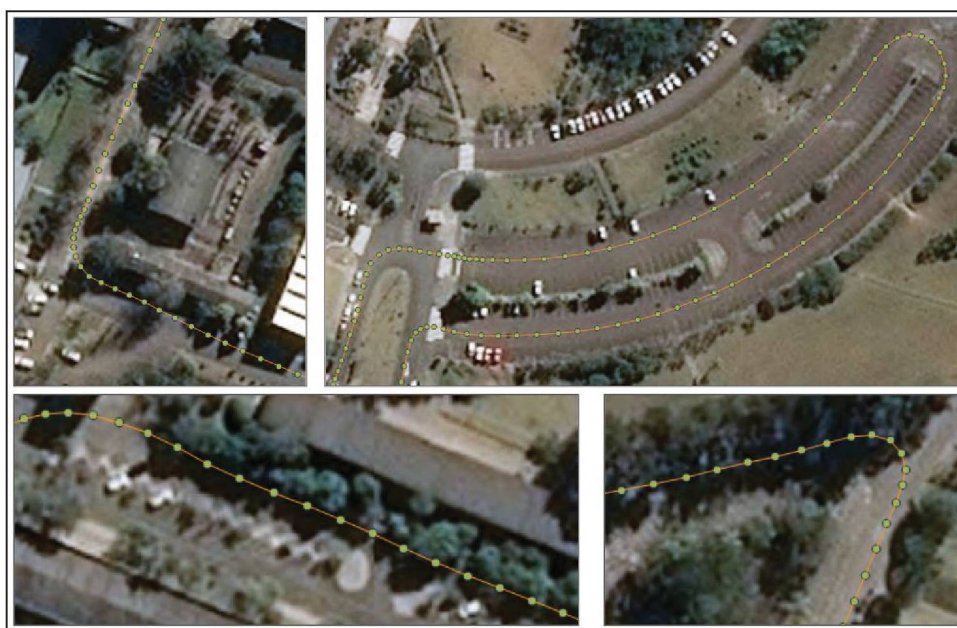
Quanto aos dados de orientação, parametrizados como quatérnios unitários (vide 2.2.1), demonstrados no capítulo 3 na forma da figura 3.6b, é perceptível que partir da inspeção visual de tal gráfico, é possível constatar os seguintes pontos:

- I. A proximidade das componentes X e Y de zero mostra a correta materialização de um sistema horizontalizado, como pode ser visto na figura 3.6a: são quilômetros de deslocamento horizontal para cerca de 20 metros de deslocamento vertical;
- II. As componentes Z e W do vetor quadri-dimensional correspondem a maior parte contribuinte para o módulo de valor fixo, e:
 - A. O predomínio da componente Z indica que o eixo de Euler é normal a um plano horizontalizado, correspondente à tangente da linha de força do campo gravitacional terrestre local, corroborando com o item I;
 - B. As mudanças na componente W indicam, como é possível visualizar e estabelecer correspondência com a figura 3.6b à mudança de azimuth da trajetória, que será o ângulo de rotação ao redor do eixo de Euler;

1. Dada a baixíssima mudança da direção do eixo de euler, o valor da componente Z muda de maneira dependente desta variação na componente W;

Com relação ao posicionamento GNSS, nenhuma metodologia numérica para avaliação da qualidade dos dados foi planejada. O único teste realizado visando ter uma noção da qualidade dos dados, está baseado na figura 4.1, no qual a trajetória foi sobreposta a ortofoto da base do *Bing Maps*, que apresenta uma resolução espacial de 30 cm⁴⁹.

Figura 4.1: Detalhes da trajetória GNSS sobreposta a ortoimagem do Bing



FONTE: O AUTOR (2019)

É possível visualmente verificar, que a trajetória é consistente com a ortoimagem de outra base. A imagem e a trajetória são: dados diferentes; com origens diferentes; e dentre outros, com fontes de erros totalmente diferentes. Assim sendo, é razoável afirmar que a consistência apresentada entre ambas, é um forte indício de que a trajetória GNSS apresenta qualidade suficiente para uma série de aplicações ligadas à estudos de vegetação urbana. Na falta de uma métrica objetiva

⁴⁹ Fonte: <<https://www.microsoft.com/en-us/maps/licensing/bing-maps-api-features-overview>>

da qualidade posicional, pode-se afirmar com segurança que resultou em melhor que 2 metros, dado o fato que é o comprimento mínimo para uma pista de asfalto.

A interpolação linear foi implementada com sucesso. Para demonstrar tal êxito (além de outras questões), foi elaborado o mapa apresentado no Apêndice 5.

4.3 ANÁLISE DOS RESULTADOS DE SEGMENTAÇÃO PELA CNN

Ademais das configurações apresentados na tabela 4.2, sobretudo modelos de CNN, cabe lembrar que outras configurações também foram testadas, no entanto para as análises foram separados apenas as melhores sequências de treinamento.

Vale lembrar que os cenários A-I e A-II são as mesmas redes neurais, sendo que a grande diferença se deu na forma da avaliação. Como pode ser observado na tabela 4.2, para A-I analisou-se o resultado comparando com imagens contendo toda a vegetação, incluindo gramíneas, ao passo que para A-II restringiu-se a apenas à vegetação não-rasteira. O motivo por trás de tal análise dicotômica é uma nuance do *dataset cityscapes*, uma vez que em suas máscaras de classificação a classe “vegetação” não possui gramíneas, que estão inseridas na classe “terreno”⁵⁰.

A metodologia para validação, segue o apresentado no capítulo 3: para cada época (uma CNN distinta), cada imagem disponível para validação é comparada com a imagem classificada pela CNN e então as métricas de erro são calculadas a partir dos somatórios dos grupos verdadeiro/falso positivo/negativo.

A escolha dos modelos de estrutura das redes e do *frontend* para ligação entre camadas foi arbitrária. Houve também grande disparidade no *Hardware* empregado para treinamento, para os cenários A, utilizou-se um computador sem placa gráfica com capacidade de uso para *Deep Learning*, ao passo que B foi treinado com computadores com processadores dedicados, disponibilizados pelo google. Isso justifica a grande disparidade existente na combinação de épocas de treinamento e tamanho de lote de processamento.

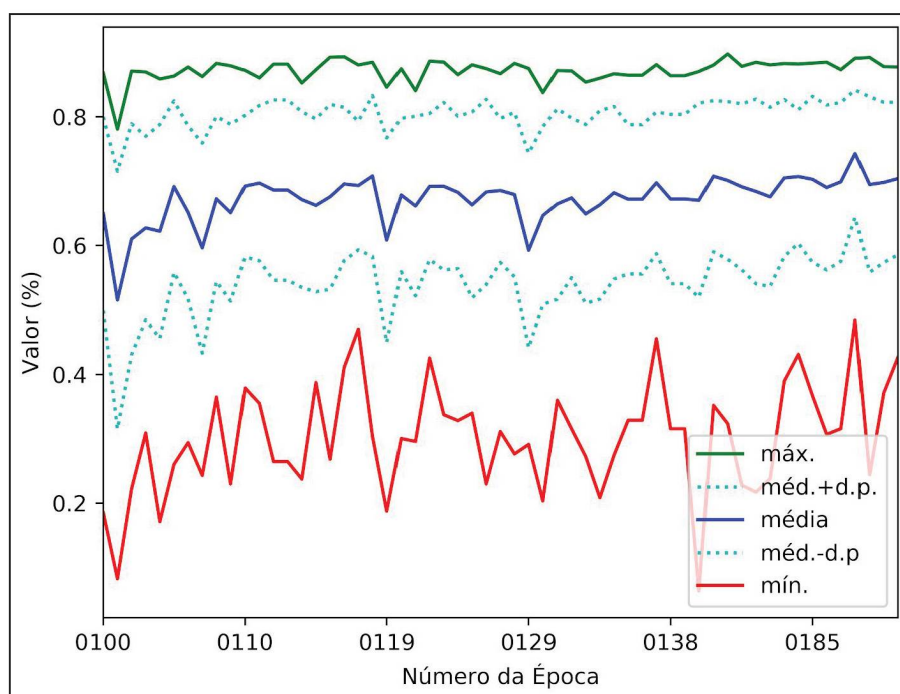
Para demonstração da evolução da rede nos cenários testados, estão apresentados os gráficos de evolução da métrica IoU nas figuras 4.2, 4.3 e 4.4,

⁵⁰ Não foi possível mesclar tal classe à vegetação para análises ou mesmo no treinamento, uma vez que ela pode englobar alguns tipos de pavimento particular, solo exposto, etc...

respectivamente para os cenários apresentados na tabela 4.2. Nas imagens estão demonstrados Os valores máximo, médio e mínimo para cada época, assim como duas linhas contendo média mais e média menos desvio-padrão, para indicativo da dispersão.

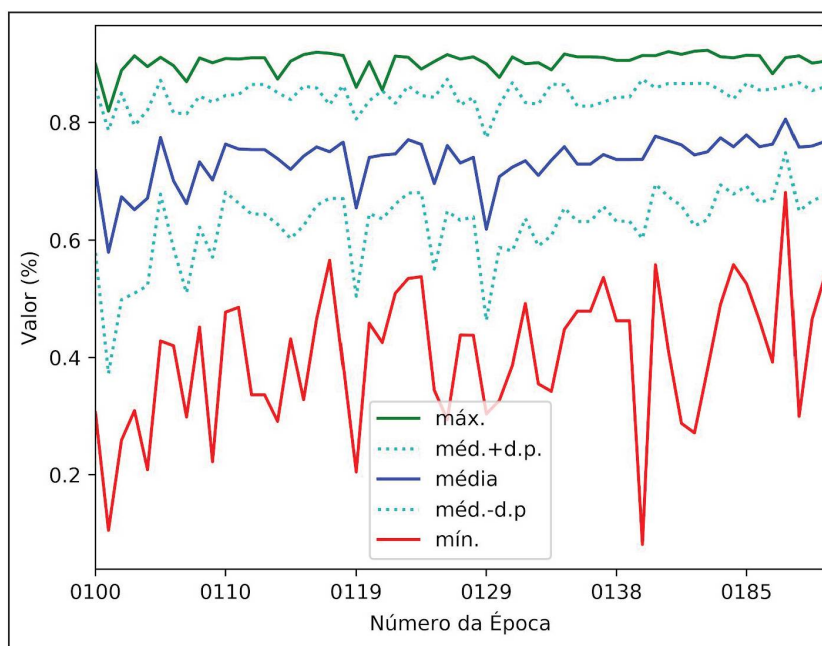
A escolha do IoU para essa demonstração se deve a duas questões: ela pode ser considerada a medida mais representativa da qualidade para os dados, por não superestimar os verdadeiros positivos, além de ter um significado geométrico para classificações binárias, onde inclusive apresenta melhor desempenho por não levar verdadeiros negativos em consideração; a segunda grande razão se deve ao fato que os gráficos das demais métricas apresentaram comportamentos globais semelhantes.

Figura 4.2: Gráfico da Evolução da Métrica IoU para o Cenário A-I



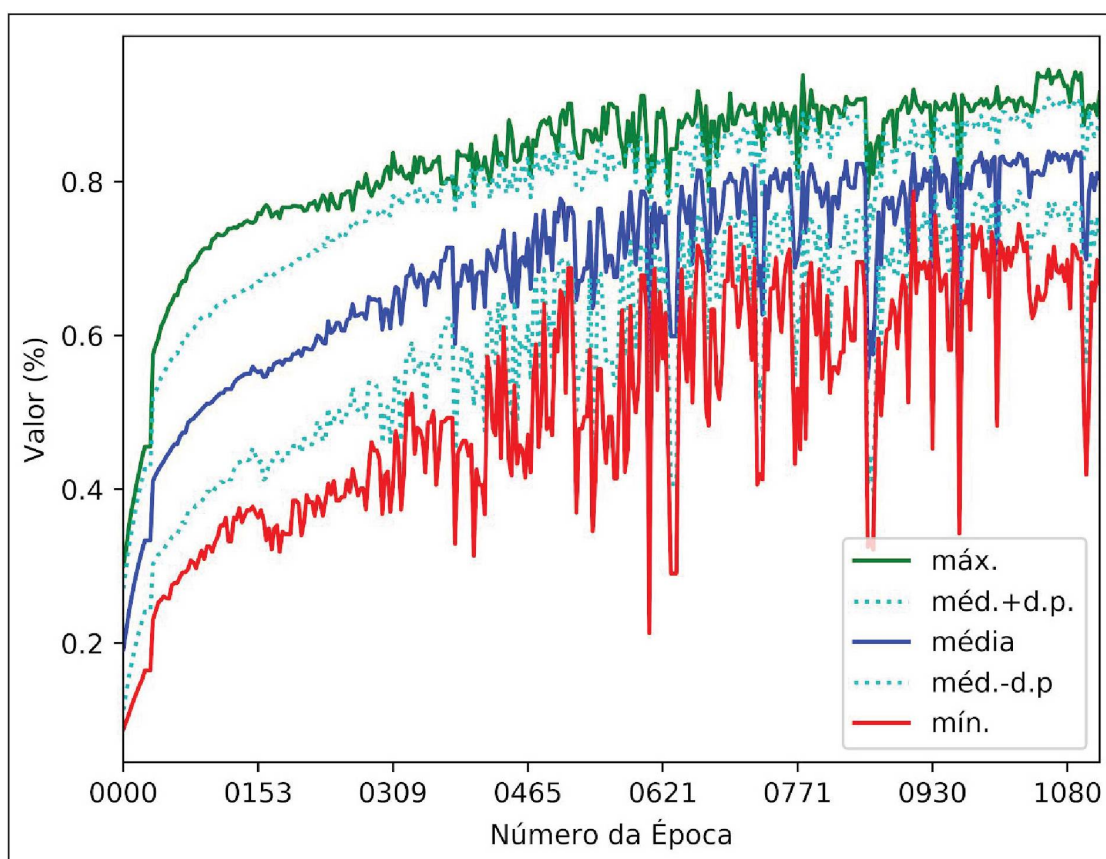
FONTE: O AUTOR (2019)

Figura 4.3: Gráfico da Evolução da Métrica IoU para o Cenário A-II



FONTE: O AUTOR (2019)

Figura 4.4: Gráfico da Evolução da Métrica IoU para o Cenário B



FONTE: O AUTOR (2019)

A primeira diferença notável entre os cenários A e o B é a densidade de pontos presente, dada a maior quantidade de épocas disponíveis para avaliação, gerando, inclusive uma aparência mais ruidosa. Na figura 4.4 pode-se acompanhar muito bem a evolução desde o começo, para o cenário B. Entretanto, não se verifica uma evolução partindo de um patamar próximo a zero nos cenários A, isso se deve ao fato que os arquivos das primeiras épocas foram perdidos, entretanto, isso não tem maiores efeitos nas análises.

A comparação de A-I com A-II, mostra que notavelmente o desempenho de A-I foi inferior, o que já era esperado, dado que o conjunto A-I foi testado com relação a redes que não foram treinadas considerando gramíneas como vegetação. De fato é possível observar um comportamento quase sistemático, de que os valores calculados para as métricas de A-I estão a um patamar fixo, o que provavelmente reflete a quantidade de vegetação rasteira existente no conjunto de dados de avaliação.

Comparando-se o gráfico de A-I com B, se percebe que os patamares para o valor médio são semelhantes, no entanto, o afastamento da média com relação aos valores extremos é maior em A-I.

Assim, uma vez calculando-se os scores médios para cada época, utilizou-se o critério apresentado no capítulo 3: definiu-se empiricamente como melhor época aquela contendo a maior média em uma maior quantidade de métricas. Assim, as melhores épocas foram, considerando a ordem de métricas como TA, VPP, Sens., F1, IoU, respectivamente:

- A-I: 200, 121, 125, 200, 200;
- A-II: 200, 121, 200, 200, 200;
- B: 984, 909, 930, 1059, 1059;

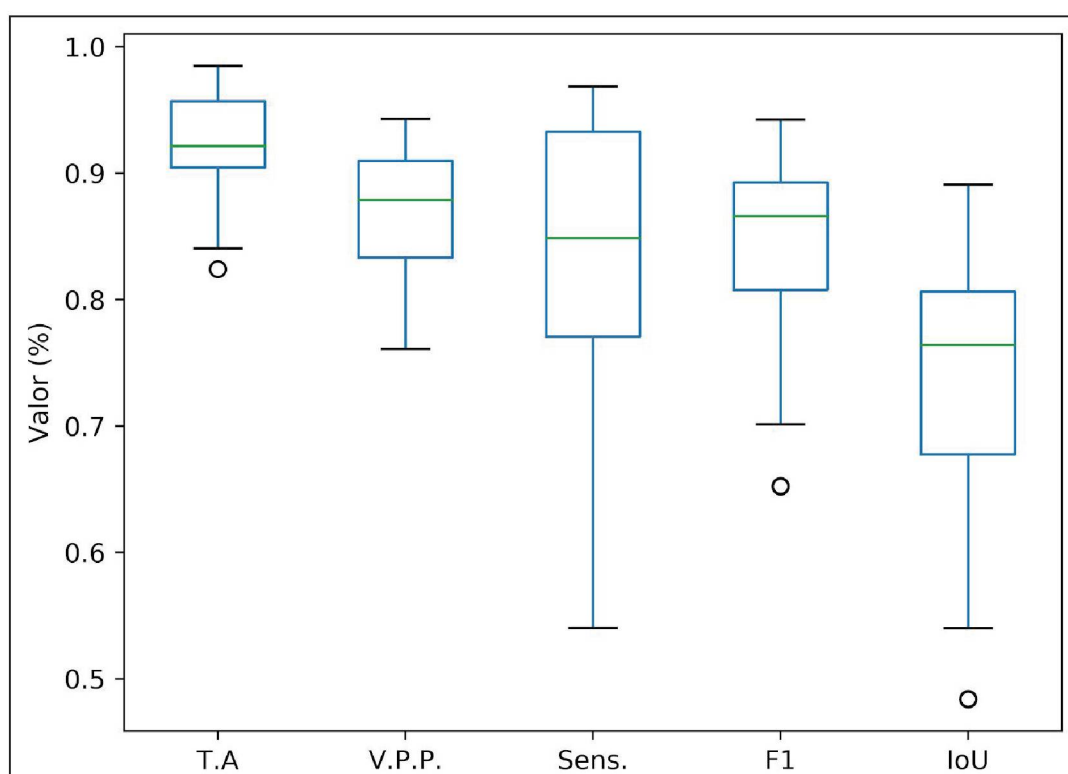
Então para os cenários A se escolheu a época 200 e para B a época 1059⁵¹. Pode-se observar que não se tratam das últimas épocas treinadas, mas sim épocas próximas, pois além das variações (podem se observar algumas bem acentuadas, sobretudo na figura 4.4) que ocorrem época a época, com o passar do tempo

⁵¹ Lembrando que não estão inseridos no mesmo sistema de contagem ordinal.

eleva-se a probabilidade de o modelo sofrer de *hiperparametrização*, tornando-se mais instável.

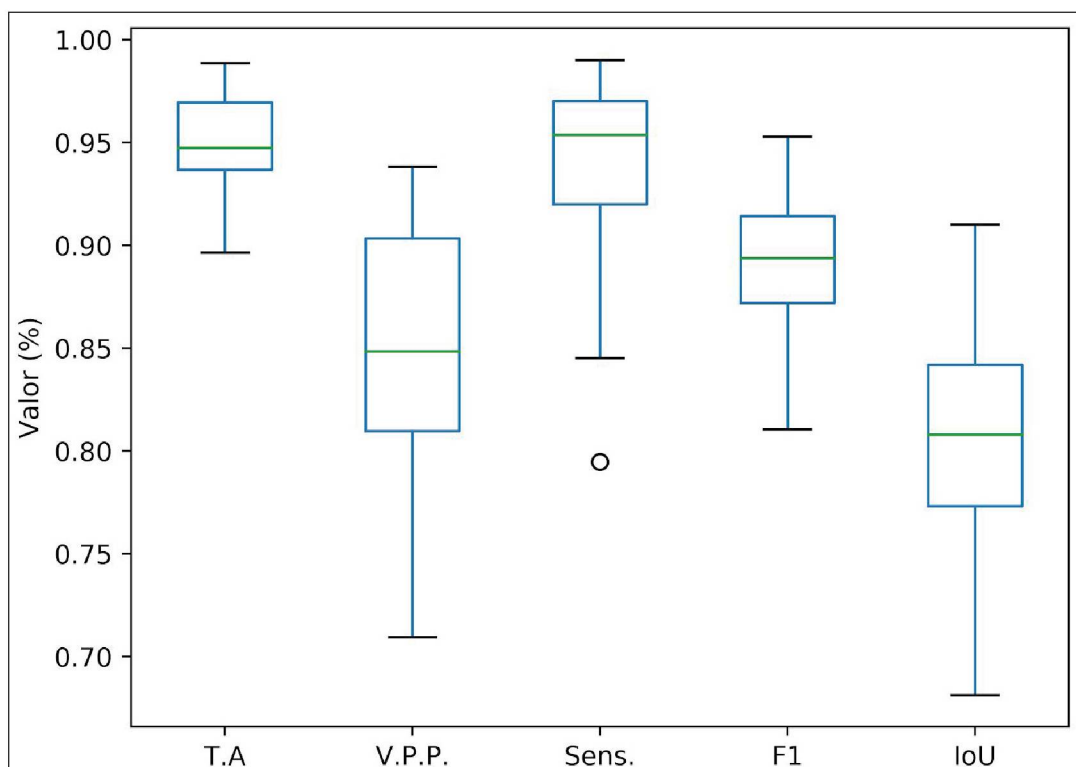
Assim, uma vez definida a melhor época, foram separadas as métricas de erro correspondentes. Para visualização dos resultados como um todo, foram gerados os *box-plots* contendo o resumo dos dados de validação para os cenários testados, sendo apresentados respectivamente aqui por meio das figuras 4.5, 4.6 e 4.7.

Figura 4.5: Box-Plots das métricas para a melhor época do cen. A-I



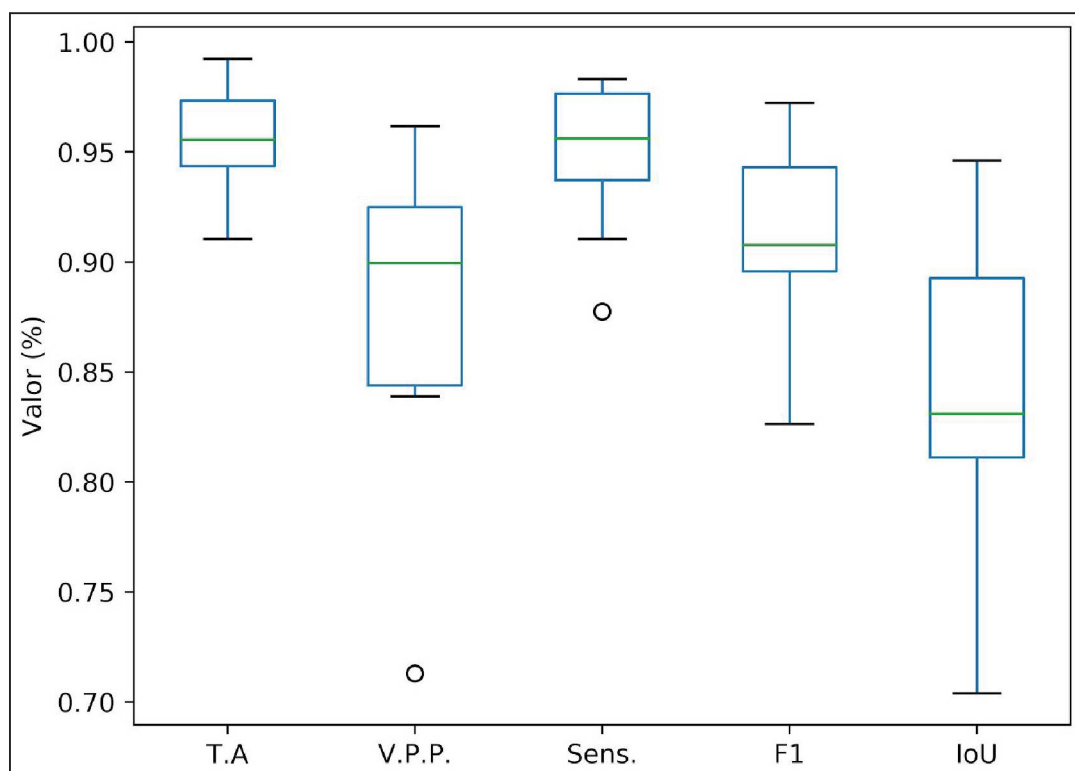
FONTE: O AUTOR (2019)

Figura 4.6: Box-Plots das métricas para a melhor época do cen. A-II



FONTE: O AUTOR (2019)

Figura 4.7: Box-Plots das métricas para a melhor época do cen. B



FONTE: O AUTOR (2019)

Percebe-se, ao se olhar para os valores numéricos das métricas, que a melhor época do cenário B se saiu melhor em quase todas as comparações, tanto em valor absoluto quanto, em dispersão. Para que o leitor tome conhecimento do valor numérico de tais métricas, na tabela 4.3 estão apresentados valores de estatística descritiva ligadas a tais métricas..

Tabela 4.3: valores de estatística descritiva da validação das melhores épocas

Cenário	Medida Estatística	T.A.	V.P.P	Sens.	F1	IoU
A-I	máx.	98.51	98.24	97.54	94.23	89.10
	mediana	92.15	93.02	87.35	86.62	76.40
	mín.	82.40	79.36	52.94	65.21	48.38
	média	92.59	92.06	84.15	84.85	74.27
	desv. pad.	3.75	3.69	11.24	6.92	9.90
	d.p. média	0.53	0.52	1.59	0.98	1.40
Cenário	Medida Estatística	T.A.	V.P.P	Sens.	F1	IoU
A-II	máx.	98.87	98.03	99.02	95.30	91.01
	mediana	94.74	92.41	95.38	89.38	80.80
	mín.	89.66	78.75	79.47	81.05	68.14
	média	95.04	91.61	93.91	89.12	80.55
	desv. pad.	2.24	3.82	4.43	3.51	5.63
	d.p. média	0.32	0.54	0.63	0.50	0.80
B	máx.	98.34	98.08	99.53	97.23	94.61
	mediana	96.03	93.50	98.73	90.77	83.10
	mín.	92.68	82.00	96.20	82.63	70.41
	média	95.71	91.55	98.46	91.15	83.97
	desv. pad.	2.04	4.59	1.04	4.13	6.88
	d.p. média	0.64	1.45	0.33	1.31	2.17

FONTE: O AUTOR (2020)

A tabela, complementa a informação apresentada nos *box-plots*. Contudo, não se pode estritamente comparar a rede escolhida para AI e A-II com a escolhida

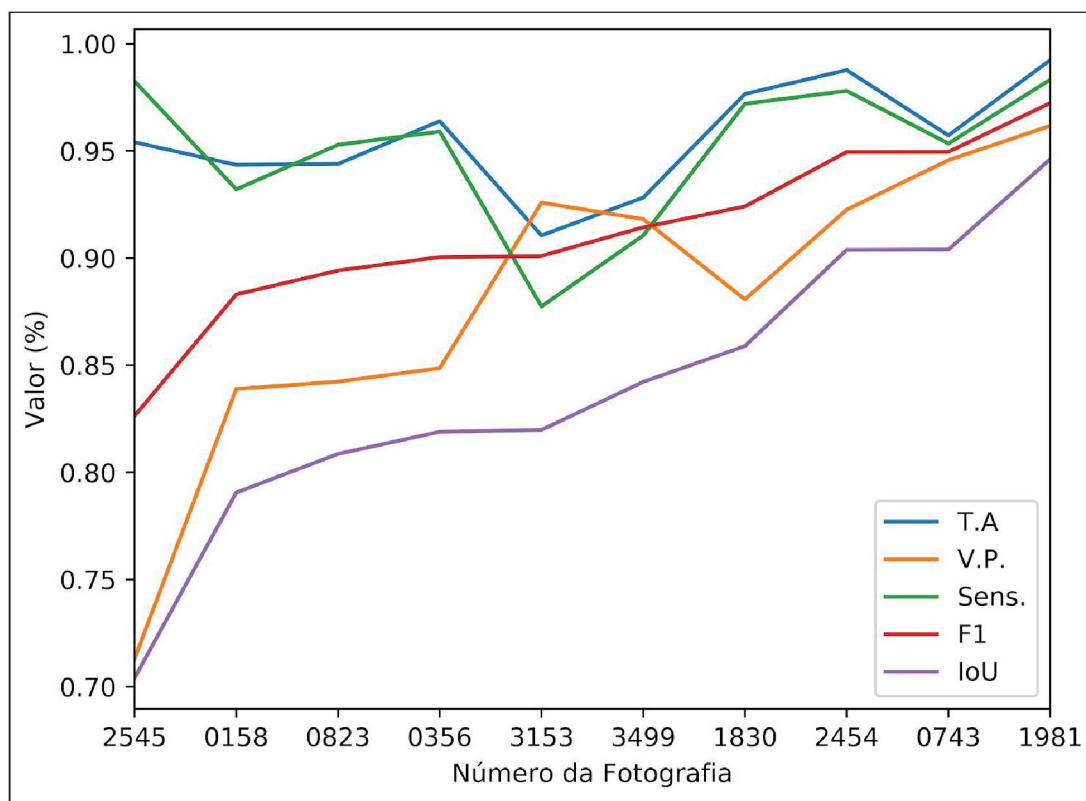
para B, uma vez que cada uma interpreta “vegetação” de maneiras diferentes, como explanado em 3.2.3.1.

Pode-se verificar como a TA é uma métrica inflada: ao se observar o valor mínimo dentre os cenários, tem-se o quase o mesmo valor do maior IoU médio. Isso poderia gerar uma interpretação errônea de “resultado satisfatório”. Tal fato ocorre sobretudo por levar em consideração os verdadeiros negativos, e é possível existir uma imagem com apenas 10% dos pixels correspondentes a vegetação, faz maior sentido avaliar apenas os pixels de interesse. No entanto, a TA ainda assim é um indicativo da eficiência da rede em evitar o oposto, os falsos positivos.

Os valores apresentados são realmente positivos, ainda que cabe destacar, que principalmente para a rede gerada apenas com dados próprios, a validação externa foi feita com apenas 10 imagens, de modo a ser uma avaliação com grande probabilidade de não ser representativa o suficiente.

A escolha da melhor rede se deu pela comparação dos valores médios (média e mediana) da métrica IoU, por ser considerada a métrica mais representativa. Assim, tem-se como eleita a rede definida pela melhor época do cenário B, com 83% de IoU médio. Na figura 4.8 estão exibidas as métricas individuais por imagem para a CNN escolhida, organizadas por ordem crescente de *score* IoU.

Figura 4.8: Métricas de erro individuais da CNN escolhida

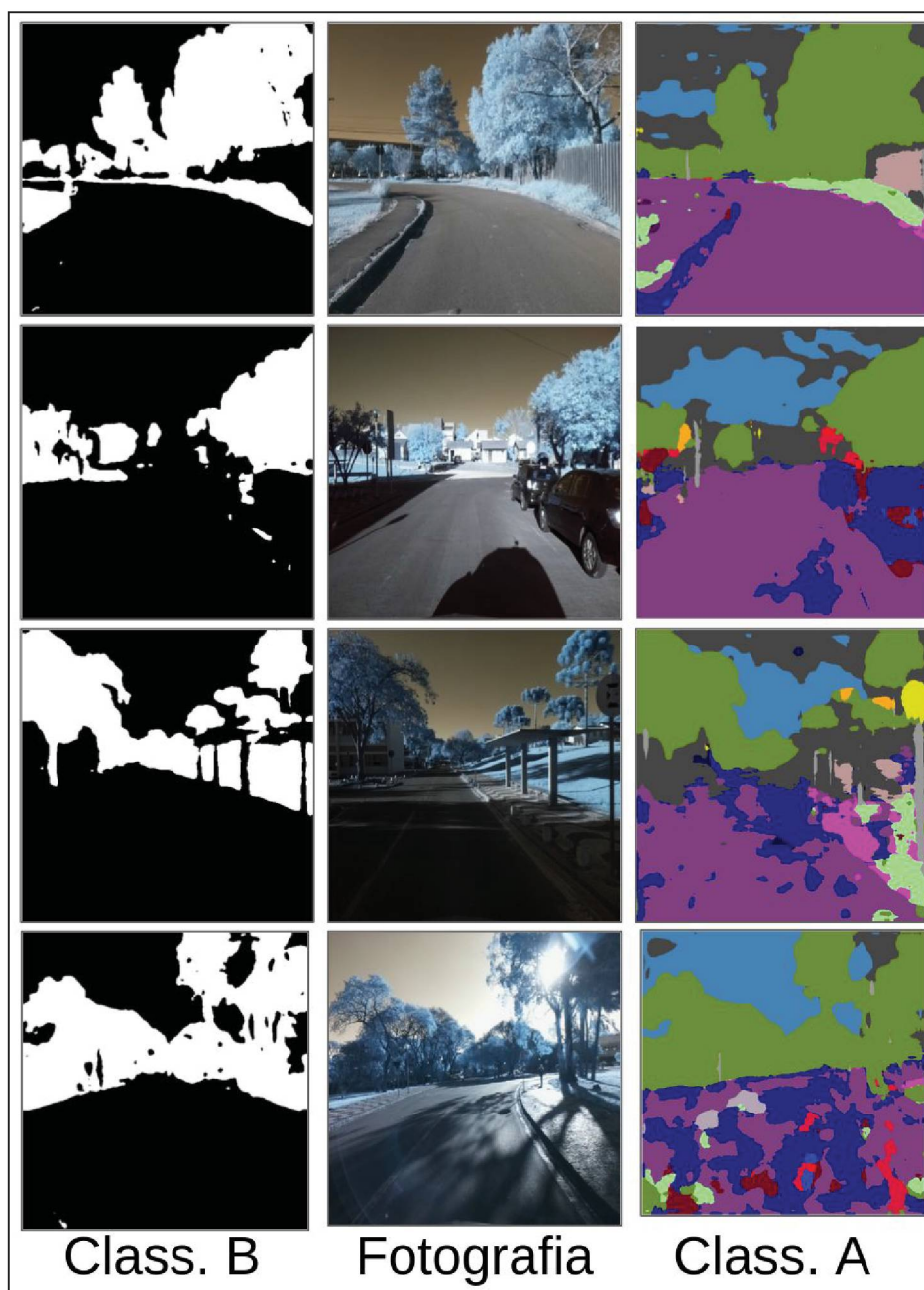


FONTE: O AUTOR (2019)

O score F1 é a média harmônica da VPP e da Sensibilidade (SE), havendo sempre uma diferença monotônica entre ele e o IoU, diretamente proporcional à diferença entre a sensibilidade e a VPP. Ao contrário da média geométrica, ela penaliza a disparidade entre suas entradas, devido a isso, pode-se observar que o seu valor mais baixo, ocorre a despeito de ser composta pelo maior valor de sensibilidade, já que incorpora também o menor valor de VPP. A diferença entre a sensibilidade e o VPP mostram que a rede se saiu melhor em evitar falsos negativos do que falsos positivos, isto é: a rede é mais eficiente em identificar o que é vegetação do que em excluir o que não é.

Uma vez definida a “melhor rede neural”, é possível gerar as máscaras de classificação para todas as imagens tomadas em um levantamento, e então combiná-las com a imagem para geração da imagem contendo apenas vegetação (e os falsos positivos residuais). Assim, como expressão do resultado, na figura 4.9 estão apresentadas amostras da classificação realizada, com as melhores épocas de cada cenário testado.

Figura 4.9: Amostras de Resultados da Classificação Realizada



FONTE: O AUTOR (2019)

Pode-se perceber que a classificação realizada por meio do cenário “B” possui melhor “detalhamento-fino”⁵², dado que nas amostras de testes considerou-se os “buracos” existentes dentro dos polígonos de vegetação, assim como contornos manualmente classificados de maneira mais detalhada.

⁵² Tradução Livre para “*fine-grained detail*”.

Recomenda-se ao leitor a visualização do resultado do processamento executado sobre uma sequência de imagens, apresentado na forma de vídeo e disponibilizado no seguinte link: <https://rebrand.ly/resultados_smmt_veg>, dado que uma sequência de imagens traz uma percepção mais precisa a respeito da qualidade da classificação realizada.

Vale lembrar que os cenários A foram treinados dispondo apenas da informação da banda R do *dataset cityscapes*, testados apenas sobre a banda R dos conjuntos de teste, de modo a haver uma clara sub-otimização no processo. Além de que sem dúvida há interferência de frequências das bandas G e B na banda R do sensor utilizado pela equipe do *Cityscapes* para captura, como demonstrado na figura 2.8, assim, não há como garantir que está sendo realizada a captura da mesma região do espectro eletromagnético. Assim, ainda que tenha performado de maneira relativamente inferior, seu resultado seguiria sendo interessante.

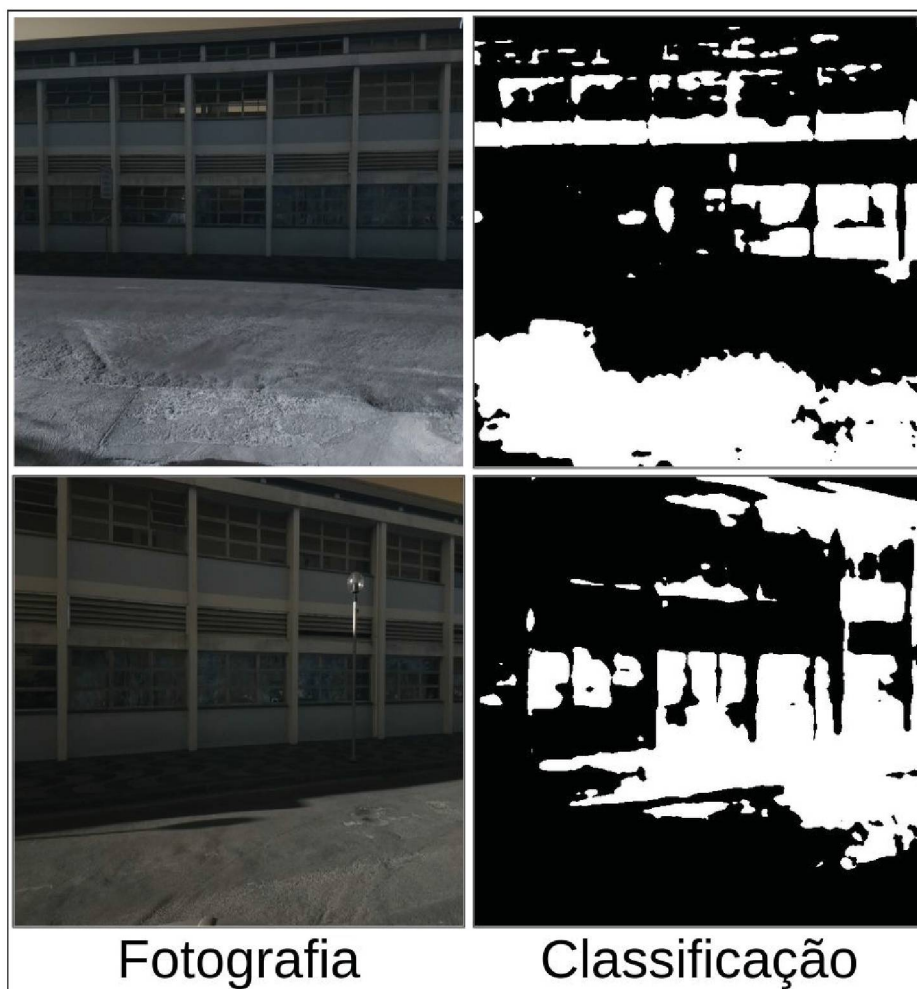
Na seção 4.3.1 será apresentada uma ponderação importante com relação às métricas obtidas para a rede definida como a “mais eficiente”.

4.3.1 Análise da classificação em Fotografias sem Vegetação

Para a rede criada no cenário B, uma das questões que pode evidenciar deficiências do modelo treinando, é a ao se testar imagens que não contém vegetação. Nesse caso, não haverá verdadeiros positivos, nem falsos negativo, uma vez que a máscara de classificação real se trata apenas de uma imagem preta. Imagens de tal tipo não foram incluídas no treinamento à época em que o mesmo foi realizado⁵³, assim como não existem no conjunto de validação. Então, executou-se a passagem de algumas imagens pertencentes a tal grupo na rede utilizada, e uma amostra do resultado pode ser vista na figura 4.10.

⁵³ E a realização de um novo treinamento com a inclusão de tais amostras já está prevista como trabalho futuro.

Figura 4.10: Demonstração do resultado gerado para a CNN Final em imagens sem vegetação



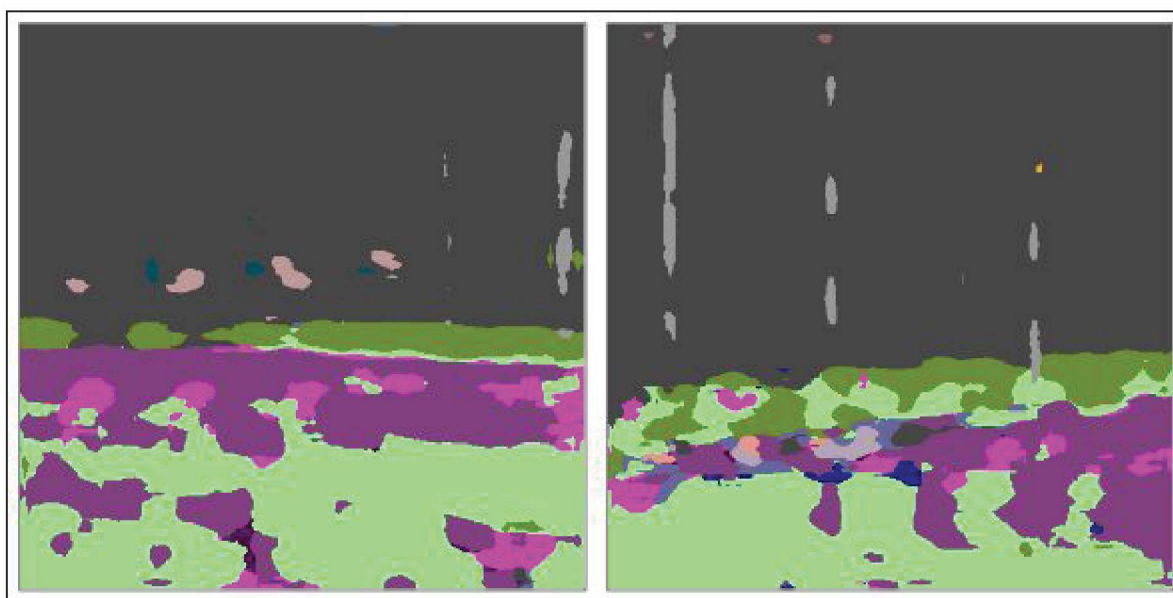
FONTE: O AUTOR (2019)

Com a falha da rede, que apresentou bom desempenho nas amostras de validação, pode-se perceber a importância de abordar todas as possibilidades existentes. A falta de amostras sem a presença de vegetação, provavelmente condicionou a construção de um modelo que pressupõe uma necessária existência de vegetação.

Pode-se perceber que na primeira imagem, a região segmentada na parte inferior, tem o formato semelhante à copada de uma árvore, já que a rede aprendeu a identificar formas. Também é perceptível que boa parte da construção é erroneamente classificada, já que a resposta espectral é semelhante à vegetação (próximo do ciano), também se assemelhando à vegetação vista de longe (perde textura).

Uma possível solução é a de realizar uma classificação prévia nas imagens (1), utilizando a melhor época do cenário A-II, e então eliminar da máscara de classificação (2) fornecida pela melhor época do cenário B a região identificada como “edificações” na primeira classificação. Na figura 4.11 estão exibidas as mesmas figuras da imagem anterior (4.10), porém dessa vez classificadas pela rede referente ao cenário A-II, onde a cor cinza-escuro correspondente à classe “edificação” e a cor verde-escuro à “vegetação”.

Figura 4.11: Demonstração do resultado gerado para a CNN Final em imagens sem vegetação



FONTE: O AUTOR (2019)

Pode-se observar duas questões principais:

1. Em parte da região erroneamente classificada nas imagens da figura 4.10, a classificação multi-classe classificou boa parte da edificação como edificação de fato, assim sendo, poderia-se sim utilizar essa outra classificação para eliminar as edificações da outra classificação;
 - a. Entretanto, na região correspondente ao asfalto, há grande presença da classe “terreno” (verde-claro), que em diversas outras imagens seria a classe atribuída à regiões de vegetação;

- b. Há outras classes, como cinza claro (poste) na região erroneamente identificada anteriormente, que também poderiam ser utilizados na exclusão proposta;
- 2. Contudo, pode-se observar também *pixels* erroneamente classificados como vegetação, de modo que não seria recomendável simplesmente substituir a classificação anterior por esta;

4.4 ANÁLISE DA VALIDADE DOS VALORES DE NDVI CALCULADOS

Uma importante etapa não realizada no presente trabalho foi a calibração radiométrica das respostas espectrais fornecidas pelo sensor, a problemática envolvida é abordada com maiores detalhes no item 6 da lista apresentada no apêndice 3. Assim sendo, aqui serão analisados os valores obtidos com as imagens registradas pelo sistema.

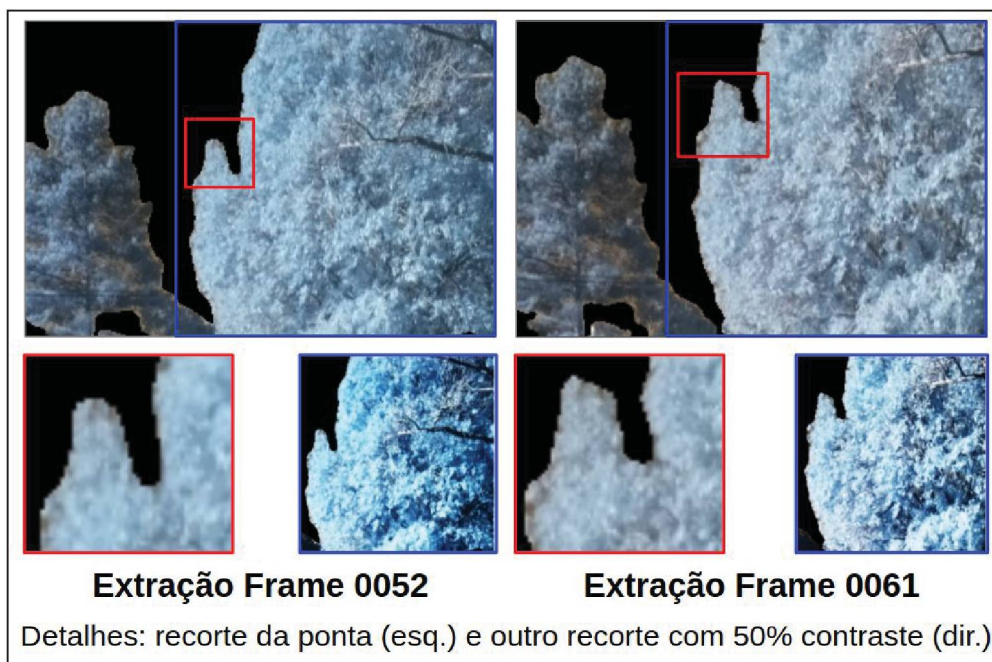
O primeiro problema a ser observado é a mudança nos tons de cinzas registrados nas diferentes bandas, para um mesmo alvo ao longo de diferentes imagens, a problemática está demonstrada na figura 4.12⁵⁴.

É perceptível que na primeira⁵⁵ imagem (0052) os pixels correspondentes à folhagem estão mais próximos da falsa-cor Ciano, ao passo que na segunda estão mais “esbranquiçados”, o detalhe com 50% de contraste foi inserido para facilitar a percepção de tal fenómeno por parte do leitor. Tal problema ocorre devido ao balanço de branco configurado como automático, que tem como função central compensar a mudança na iluminação recebida, de modo a manter o nível de exposição uniforme. Pode-se perceber que o balanço de branco automático permaneceu constante para a banda NIR, ao passo que aumentou para a banda R, já que o nível de exposição ficou muito próximo e a tonalidade não mudou de matiz, apenas foi des-saturada.

⁵⁴ Recomenda-se a leitura em versão digital para correta visualização da mesma.

⁵⁵ Aqui a ordinalidade refere-se estritamente à sequencialidade na tomada das fotografias.

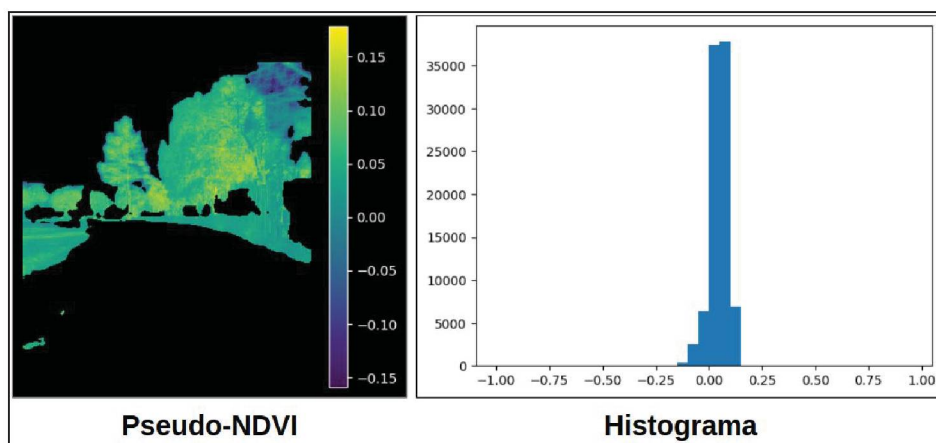
Figura 4.12: Mudança nos tons de cinza nas mesmas feições, imagens diferentes



FONTE: O AUTOR (2019)

Este fator por si só já afetaria a confiabilidade dos valores de NDVI, entretanto, também há um problema com a escala dos valores calculados pela simples aplicação inadvertida da fórmula canônica⁵⁶ de tal índice. Para demonstração de tal problemática, abaixo, na figura 4.13 estão demonstrados o histograma e o recorte com a aplicação de uma escala de cores de um típico resultado advindo do conjunto de dados de teste.

Figura 4.13: histograma e imagem com valores típicos de NDVI calculados

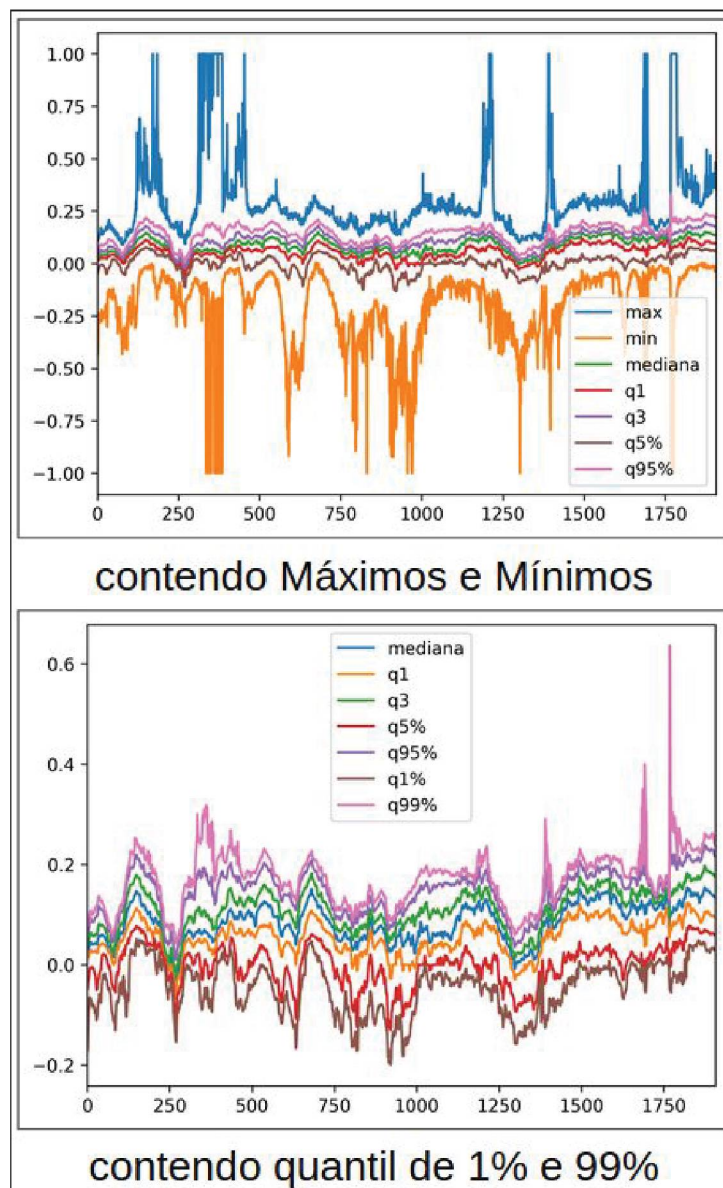


FONTE: O AUTOR (2019)

⁵⁶ De fato poderia se propor uma versão adaptada para as particularidades radiométricas do imageador empregado.

Foram extraídas as informações de máximo, mínimo e alguns quantis ao longo de cerca de 1900 imagens sequenciais, para visualização do fenômeno de maneira mais global. Os gráficos estão apresentados na figura 4.14.

Figura 4.14: variação dos valores calculados de NDVI



FONTE: O AUTOR (2019)

Observando as ilustrações, é possível perceber que os valores não se encontram no intervalo correto indicado na literatura (correspondente a $\{-1,1\}$), uma vez que ademais dos valores extremos⁵⁷, os valores flutuam próximos a zero,

⁵⁷ Que são *outliers*, de objetos não pertencentes a vegetação (Falsos Positivos), que

entretanto há grande presença de vegetação saudável, de modo que altos valores de NDVI deveriam ser frequentes, todavia, não é o que ocorre. Os valores médios para os quantis extremos são: -0,000602 para 5%; 0,145 para 95%; -0,0464 para 1%; e 0,171 para 99%. Isto é, no extremo inferior o valor mais provável está próximo de -0.05 e na outra ponta próximo a 0.15 (considera-se 90% dos pixels classificados como vegetação). Assim sendo, pode-se verificar que os valores de NDVI calculados para o presente trabalho não são confiáveis, fazendo-se necessária a realização de uma calibração radiométrica.

Todavia, a presença de respostas espectrais relativamente uniformes ao longo do dataset de recortes, fala em favor da qualidade da classificação, uma vez que para quase todos os casos, os pixels com respostas espectrais totalmente dissonantes são menos de 2%, como é visível nos gráficos da figura 4.14.

5 CONSIDERAÇÕES FINAIS

O presente trabalho apresentou dois temas relevantes para a área de Ciências Geodésicas: o desenvolvimento de um SMMT cujos dados sejam propícios para mapeamento da vegetação urbana; e a classificação das imagens obtidas pelo sistema através da aplicação de uma CNN treinada por meio do uso de *Deep Learning*, etapa considerada como predecessora a qualquer análise a ser realizada com os dados coletados pelo sistema proposto.

Primeiramente, alcançou-se o sucesso na montagem de uma plataforma estável e de baixo custo, capaz de gravar todos os dados em um mesmo sistema de tempo, sendo um sistema de fácil e remota operação. Com relação aos dados de navegação, ainda que não tenham sido realizados testes mais introspectivos e cuidadosos, há indicativos de sua estabilidade e de que atenda à faixa de precisão adequada para vários tipos de estudo. Há ainda uma série de melhorias a serem feitas, abordadas no apêndice 3, dentre elas, destaca-se a calibração radiométrica da câmera.

Os resultados obtidos para a classificação das imagens são muito positivos, ainda que importantes ressalvas devam ser feitas: a rede não se sai bem na ausência de vegetação; e mais amostras de testes são necessárias para uma avaliação não-tendenciosa. Cabe lembrar ao leitor que o processo de classificar uma imagem manualmente toma cerca de 2 horas para uma classificação *booleana*, assim sendo, os resultados utilizando apenas os dados do *dataset cityscapes* também devem ser levados em consideração.

A capacidade das CNNs de gerar de modelos complexos para segmentações arbitrárias, incorporando tanto aspectos geométricos como radiométricos, traz uma quebra de paradigma para tal tópico, possibilitando modelos cada vez mais acurados.

O sistema proposto possui os quesitos para aplicação em larga escala. O baixo custo envolvido viabiliza a instalação de unidades em uma grande quantidade de veículos que poderão atuar como agentes de mapeamento contínuo, sobretudo caso tais veículos pertençam à frotas de mobilidade urbana.

Os dados coletados, têm grande potencial para uso em estudos da saúde e da presença da vegetação em cenários urbanos, entretanto, cabe a lembrar que para a realização de inferência a partir dos mesmos, se faz necessária uma abordagem multidisciplinar, envolvendo uma equipe contendo profissionais da área das ciências biológicas e/ou de engenharia florestal ou similares. O autor do presente trabalho no momento de sua entrega possui conhecimentos aquém dos necessários para se propor a realizar tais análises. Todavia, ao leitor interessado, é incentivado o estabelecimento de contato para intercâmbio de conhecimento.

Assim, com a realização do presente trabalho, processamento dos dados de amostra, geração dos resultados e análises realizadas sobre os mesmos, pode-se concluir que os objetivos apresentados no capítulo 1 foram alcançados.

5.1 SUGESTÕES PARA TRABALHOS FUTUROS

1. Uso de computadores com maior capacidade de processamento: *Hardware* trata-se de um recurso indispensável no treinamento de CNNs, sobretudo com relação a placas de vídeo.
2. Re-treinamento da rede neural gerada contendo imagens sem a presença de vegetação, conforme mostrado no capítulo 4 a ausência de tais imagens no treinamento resultou em uma rede que não lida bem com a possibilidade de existir uma cena com ausência de vegetação;
3. Treinamento de redes, para identificação de:
 - a. Região sombreada e iluminação solar direta;
 - b. Caules/Troncos/Raízes e Folhagem;
 - c. Patologias específicas em árvores, é possível empregar CNNs para detectar elementos que indiquem possíveis problemas em indivíduos arbóreos;
 - d. Riscos elétricos concomitantes à árvores;
4. Mensuração individual de cada árvore, isto é: realização de um mapeamento de fato;
 - a. Contagem do número de árvores por imagem por meio da contagem da quantidade de caules;

- b. Cômputo das coordenadas tridimensionais para uma aproximação do ponto-base de cada árvore;
- 5. Teste de diferentes filtros para obtenção de imagens com características radiométricas diferentes: no presente trabalho se utilizou um filtro de barreira, vermelho, capaz de bloquear frequências abaixo do verde;
 - a. Uma possibilidade é o emprego de um filtro azul, que se trata de um filtro de intervalo, bloqueando verdes e vermelhos, gerando uma separação mais consistente entre a banda visível empregada e o infravermelho próximo. Entretanto, a resposta espectral da vegetação na banda do azul é diferente da vermelha, tornando-se menos distintiva para alguns alvos⁵⁸;
- 6. Uso de sensores de mudança de iluminação solar: adaptar parâmetros de exposição fotográfica e ganhos (balanços de branco) baseados em medidas;
- 7. Melhorar sub-otimizações que ocorreram ao longo do treinamento das redes, tais quais:
 - a. Em uma rede com imagens monocromáticas (cenários A, expostos no capítulo 4), realizou-se o treinamento tendo como entrada a estrutura padrão de imagem, contendo 3 camadas, assim, todos os cálculos da rede foram realizados 3 vezes;
 - b. Uso de estrutura multiclasse para uma classificação binária: buscar uma implementação otimizada para duas classes;
- 8. Uma vez que várias épocas foram treinadas e estas desempenham diferente entre si (de fato cada época é uma rede diferente), seria possível separar as N melhores redes (por exemplo, 100) e fazer um sistema de “votação”:
 - a. O pixel pertencerá, em uma classificação binária à classe que pertencer em mais de 50% das versões; ou
 - b. Será considerado apenas caso mais de 60% das classificações aponte para uma classe, entre 40% e 60%, o pixel será marcado como “em dúvida”; ou
 - c. podem ser usados outros percentuais, ou outros critérios de votação; ou a combinação de diferentes modelos de CNN;

⁵⁸ Para maiores detalhes na discussão vide HORNING (2013).

9. Incorporar restrições⁵⁹ geométricas à classificação:

- a. As máscaras de classificação precisam ter consistência estereoscópica (geometria epipolar);
- b. Para algumas feições a serem classificadas como não-árvores (caso de emprego de uma rede multi-classes), as imagens monoscópicas dos objetos jamais assumirão algumas formas (ex: uma pessoa jamais gerará uma imagem triangular); a recíproca é verdadeira também (ex: um poste com luminária esférica, sempre gerará uma imagem circular).

⁵⁹ Ao leitor interessado, recomenda-se: <https://www.tensorflow.org/neural_structured_learning/>. Ou a temática “Constrained Deep Learning”.

REFERÊNCIAS

ALHASHIMI, Anas. **Statistical Sensor Calibration Algorithms**. 2018. Doctoral Thesis. Lulea University of Technology.

BRASOLIN, S. **Quedas de árvores: o bem-estar sob risco nas cidades**. Fonte: < <https://www.ecodebate.com.br/2015/02/27/quedas-de-arvores-o-bem-estar-sob-risco-nas-cidades-artigo-de-sergio-brazolin/> >. Acesso em: 15 ago 2019.

BOYD, J. **Japan's Upgraded Mobile Mapping Technology Aims to Make Autonomous Driving Safer**. IEEE Spectrum, 2016. Disponível em: < <https://spectrum.ieee.org/cars-that-think/transportation/self-driving/japans-upgraded-mobile-mapping-technology-aims-to-make-autonomous-driving-safer> >. Acesso em: jan 2020.

BADRINARAYANAN, V., et. al. **Segnet: A deep convolutional encoder-decoder architecture for image segmentation**. IEEE transactions on pattern analysis and machine intelligence, v. 39, n. 12, p. 2481-2495, 2017.

CARTER, G. A.; KNAPP, A. K. **Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration**. American journal of botany, v. 88, n. 4, p. 677-684, 2001.

CORDTS, M., et al. **The cityscapes dataset for semantic urban scene understanding**. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

COLLINS, George W. et al. The foundations of celestial mechanics. **The Foundations of Celestial Mechanics**. Pachart Publishing House (Pachart Astronomy and Astrophysics Series. Volume 16), 1989, 158 p., 1989.

CHAKRAVARTY, S. **Here are top mobile mapping systems that would interest you**. 2018. Disponível em: < <https://www.geospatialworld.net/blogs/here-are-five-mobile-mapping-systems-that-would-interest-you/> >. Acesso em: jan 2020.

CHEN, Liang-Chieh et al. **Semantic image segmentation with deep convolutional nets and fully connected crfs**. arXiv preprint arXiv:1412.7062, 2014.

DENG, Li et al. **Deep learning: methods and applications**. Foundations and Trends in Signal Processing, v. 7, n. 3–4, p. 197-387, 2014.

DIEBEL, James. **Representing attitude: Euler angles, unit quaternions, and rotation vectors**. Matrix, v. 58, n. 15-16, p. 1-35, 2006.

EL-SHEIMY, Naser. **An overview of mobile mapping systems**. In: FIG Working Week. 2005. p. 16-21.

FAWCETT, T. **An introduction to ROC analysis.** *Pattern recognition letters* 27, no. 8 (2006): 861-874.

HARTLEY, Richard; ZISSERMAN, Andrew. **Multiple view geometry in computer vision.** Cambridge university press, 2003.

HORNING, N. **Why a red filter should work well for NDVI.** Public Lab, 2013. Disponível em: <<https://publiclab.org/notes/nedhorning/11-01-2013/why-a-red-filter-should-work-well-for-ndvi>>. Acesso em: jun 2019.

HE, K. et. al. **Deep residual learning for image recognition.** In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

KALNINS L. **Coordinate Systems.** March 2009. Disponível em: <<https://www.earth.ox.ac.uk/~larak/MMES/CoordinateSystems.pdf>>. Acesso em 21 set 2017.

KARPATHY, A., et al. **CS231N: convolutional neural networks for visual recognition.** Neural networks, v. 1, 2016. Disponível em: <<http://cs231n.github.io>>. Acesso em: 3 jan 2020.

KHAN ACADEMY. **Luz: ondas eletromagnéticas, espectro eletromagnético e fótons.** 2019. Disponível em: <<https://pt.khanacademy.org/science/physics/light-waves/introduction-to-light-waves/a/light-and-the-electromagnetic-spectrum>>. Acesso em: nov 2019.

JEFFREY, Charles. An introduction to GNSS: GPS, GLONASS, Galileo and other global navigation satellite systems. **Eng., Novatel Inc**, 2015.

JIANG, Z., et al. **Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction.** Remote sensing of environment, v. 101, n. 3, p. 366-378, 2006.

LIN, Guosheng et al. **Refinenet: Multi-path refinement networks for high-resolution semantic segmentation.** In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1925-1934.

LONG, J., et al. **Fully convolutional networks for semantic segmentation.** In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3431-3440.

LUCID VISION LABS. **Understanding The Digital Image Sensor.** Disponível em: <<https://thinklucid.com/tech-briefs/understanding-digital-image-sensors/>>. Acesso em: 3 jan 2020.

MADEIRA, S.; GONÇALVES, J.A.; BASTOS, L. **Sensor integration in a low cost land mobile mapping system.** Sensors, v. 12, n. 3, p. 2935-2953, 2012.

MASIULIONIS, T. **Automatic assessment methods of A pilot's flight in the traffic circuit.** Aviation, v. 20, n. 2, p. 53-59, 2016.

MCHUGH, S. **Understanding photography: volume 1: core concepts and techniques with digital image capture.** 2017, Cambridge In Colour.

MIKOLAJCZYK A., GROCHOWSKI M. **Data augmentation for improving deep learning in image classification problem.** 2018 *International Interdisciplinary PhD Workshop (IIPhDW)*, Swinoujście, 2018, pp. 117-122.

MILLS, D.L. **Ntp architecture, protocol and algorithms.** *Technical report*, 2002. .
Fonte: < <https://www.eecis.udel.edu/~mills/database/brief/arch/arch.pdf> >. Acesso em: 22 out 2019.

MOSTAJABI, M., et al. **Feedforward semantic segmentation with zoom-out features.** In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3376-3385.

NAVIPEDIA. **GNSS.** Disponível em: <<https://gssc.esa.int/navipedia/index.php/GNSS>>. Acesso em 21 set 2017.

NICODEMO, M; PRIMAVESI, O. **Por que manter árvores na área urbana?** Embrapa Pecuária Sudeste-Docmentos (INFOTECA-E), 2009.

NIJLAND, W. et al. **Monitoring plant condition and phenology using infrared sensitive consumer grade digital cameras.** Agricultural and Forest Meteorology, v. 184, p. 98-106, 2014.

PALAIS, B., PALAIS, R. **Euler's fixed point theorem: The axis of a rotation.** Journal of fixed point theory and applications, v. 2, n. 2, p. 215-220, 2007.

RAHMAN, M., YANG W. **Optimizing intersection-over-union in deep neural networks for image segmentation.** In *International symposium on visual computing*, pp. 234-244. Springer, Cham, 2016.

RICHARDSON T, LEVINE J. **The remote framebuffer protocol.** 2011. Disponível em : <<https://tools.ietf.org/html/rfc6143>>. Acesso em 22 set. 2018.

SHARDA, A. **How GPS Works.** Disponível em: <<https://medium.com/@aryamansharda/how-gps-actually-works-e6e0d126d2d5>>. Acesso em: nov 2018.

SILVA, K., et. al. **Hardware para controle avançado de veículo aéreo não tripulado do tipo quadricóptero.** Horizonte Científico, v. 8, n. 1, 2014.

SIORDIA, Oscar S. et al. **Wireless in-vehicle complaint driver environment recorder.** In: Proceedings of the International Conference on Signal Processing and Multimedia Applications. IEEE, 2011. p. 1-7.

SEIFF, G. **Semantic Segmentation Suite in TensorFlow**. Disponível em: <<https://github.com/GeorgeSeif/Semantic-Segmentation-Suite>>. Acesso em: Agosto de 2019.

STANSELL, T. A. **The Transit Booklet**. Disponível em: <<https://www.ion.org/museum/files/TransitBooklet.pdf>>. Acesso em 21 set 2019.

SZEGEDY, C., et. al. **Going deeper with convolutions**. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 1-9.

TALAYA, J.; et al. **GEOVAN: The mobile mapping system from the ICC**. In: Proceedings of the 4th International Symposium on Mobile Mapping Technology (MMT'2004), Kunming. 2004.

TAZARTES, Daniel. An historical perspective on inertial navigation systems. In: **Inertial Sensors and Systems (ISISS), 2014 International Symposium on**. IEEE, 2014. p. 1-5.

TOTH, C., et. al. **MOVING TOWARD REAL-TIME MOBILE MAPPING: AUTONOMOUS VEHICLE NAVIGATION**. 2007. The 5th International Symposium on Mobile Mapping Technology. Padua, Italy.

VERHOEVEN, Geert. **The reflection of two fields: electromagnetic radiation and its role in (aerial) imaging**. AARGNEWS, v. 55, p. 13-18, 2017.

VESTENA, K. **Desenvolvimento de um Sistema de Mapeamento Móvel Terrestre**. Universidade Federal do Paraná, 2017. Trabalho de Conclusão de Curso. Disponível em: <http://rebrand.ly/tcc_kaue>.

WEIER, J.; HERRING, D. **Measuring vegetation (ndvi & evi)**. NASA Earth Observatory, v. 20, 2000.

WOODMAN, Oliver J. **An introduction to inertial navigation**. University of Cambridge, Computer Laboratory, 2007.

YIN, K. **Overcome Overfitting During Instance Segmentation with Mask-RCNN**. 2019. Disponível em: <<https://towardsdatascience.com/overcome-overfitting-during-instance-segmentation-with-mask-rcnn-32db91f400bc>>. Acesso em: 3 jan 2020.

YUAN, Yuhui; CHEN, Xilin; WANG, Jingdong. **Object-Contextual Representations for Semantic Segmentation**. arXiv preprint arXiv:1909.11065, 2019.

ZHAO, Hengshuang et al. **Pyramid scene parsing network**. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 2881-2890.

ZHENG, S et al. **Conditional random fields as recurrent neural networks**. In: Proceedings of the IEEE international conference on computer vision. 2015. p. 1529-1537.

ZHOU, K., et al. **Assessing the spectral properties of sunlit and shaded components in rice canopies with near-ground imaging spectroscopy data.** Sensors, v. 17, n. 3, p. 578, 2017.

APÊNDICE 1 - OUTRAS POTENCIALIDADES DO SISTEMA PROPOSTO

Considerando que os dados coletados pelo sistema apresentado carregam um conjunto de finalidades advindas da sua própria proposição, apenas resultados diretamente ligados a tal leque de possibilidades (estudos de vegetação urbana) estão apresentados no corpo central desta dissertação.

Entretanto, os dados coletados pelo sistema podem ser empregados para outras aplicações, assim sendo, o presente apêndice visa demonstrar exemplos de tais aplicações e suas potencialidades. Assim sendo, serão apresentadas na sequência as seguintes aplicações:

1. Geração de ortofotos aproximadas a partir de cada foto individual;
2. Geração de trajetória, por odometria visual;
3. Geração de nuvem de pontos 3D colorida em falsa-cor;

Na aplicação 1, por meio de algumas suposições, sabendo o tamanho do sensor, a distância focal nominal, altura da antena com relação à plataforma e a altura do veículo utilizado, é possível obter uma aproximação das coordenadas X e Y que seriam projetadas em um plano com Z constante, perpendicular ao plano de projeção da imagem, representando o plano em que se situam as rodas do veículo, que corresponderia a um chão sem irregularidades, o fato é que para boa parte do cenário urbano, em que o veículo percorre o asfalto, essa é uma aproximação razoável, exceto em travessias elevadas ou redutores de velocidade.

Assim sendo, ao estabelecer vetorialmente uma equação para se projetar um ponto do plano da imagem no “plano do chão”, é possível projetar 4 pixels que correspondem a um retângulo na imagem e a um trapézio isósceles no outro plano. Assim, de posse dos pontos correspondentes em ambos os sistemas, é possível determinar por meio dos mínimos quadrados uma transformação projetiva, que será capaz de transformar de um plano para o outro.

Embutida na transformação está a unidade empregada para as coordenadas de destino, de acordo com o planejado, cada pixel equivalerá a um centímetro no plano projetado, isto é, a imagem resultante, com cerca de 1400 x 800 pixels,

representará uma região de 14 x 8 metros. Na figura A1.1 estão apresentados pares contendo as imagens originais e a parte retificada (correspondentes, no teste realizado à 35% dos pixels da imagem, a partir da parte de baixo⁶⁰).

Figura A1.1: exemplos de aproximações monoscópicas de ortofotos geradas com as imagens



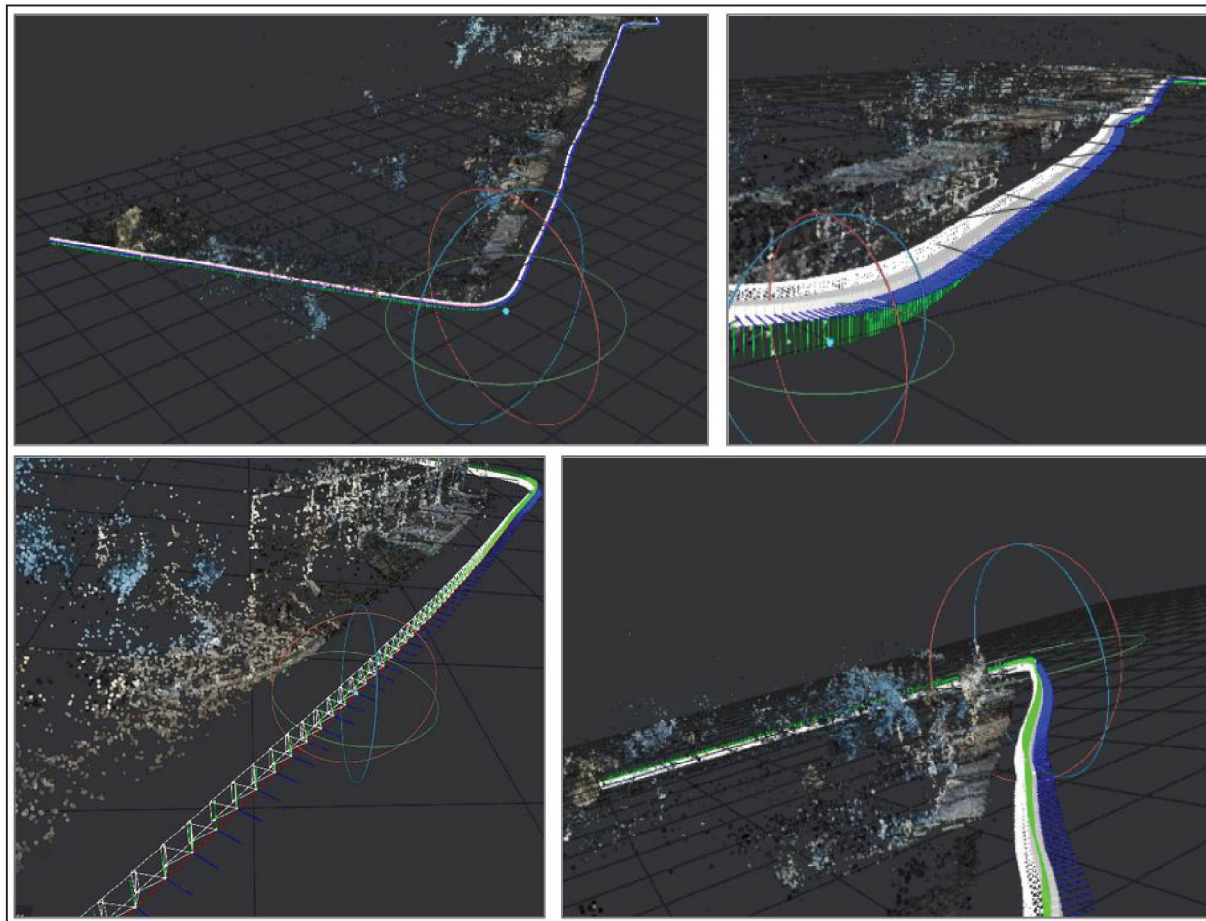
FONTE: O AUTOR (2019)

Quanto à aplicação 2, com uso do software *Meshroom*, o mesmo a ser empregado para a geração da aplicação 3, é possível gerar a trajetória percorrida pela câmera na plataforma. Considerando que na configuração empregada ao longo do presente trabalho, as imagens foram tomadas com baixa resolução e frequência relativamente alta, tais dados podem ser utilizados para tal finalidade. Na figura A1.2

⁶⁰ O Fato é que não se pode retificar a imagem inteira, uma vez que os pontos no horizonte da imagem (linha do meio) estão no infinito de qualquer plano perpendicular ao sensor que não contenha tal linha. Quanto mais próximo do centro, menor será a resolução das feições retificadas, empiricamente se chegou a 35%.

está exibida uma trajetória calculada de tal maneira. Vale lembrar que para tal teste, os dados foram coletados com a câmera instalada com visada lateral e que tal trajetória foi gerada com escala arbitrária, no sistema de eixos da primeira fotografia.

Figura A1.2: detalhes de uma trajetória gerada com as imagens do SMMT



FONTE: O AUTOR (2019)

Quanto à aplicação 3, alguns resultados de nuvens esparsas de nuvens de pontos podem ser observadas. No futuro serão realizados testes para geração de nuvens densas.

APÊNDICE 2 - OBTENÇÃO SIMPLES DE MEDIDAS DE ÁRVORES

Como já comentado no Capítulo 1, é possível apenas com uma foto, sem qualquer informação do sensor, obter medidas lineares, sendo que sua utilidade já foi comentada em 1.3, o presente apêndice será dedicado a mostrar como tais medidas podem ser obtidas a partir das imagens tomadas pelo sistema apresentado nesta dissertação.

Cabe a ressalva de que são feitas uma série de suposições a respeito das imagens e de suas características geométricas, sem as quais, os procedimentos aqui apresentados seriam totalmente inválidos. Vale lembrar também que com as fotos é possível se realizar qualquer medida em 3D, entretanto, são processos que exigem o conhecimento de uma quantidade bem maior de dados para geração e demandam maior esforço.

Na figura A2.1 está apresentado um exemplo de imagem tomada com o SMMT, cujas particularidades geométricas advindas da posição com que são tomadas⁶¹, estão aqui mencionadas:

1. O movimento da plataforma é frontal, e o plano de projeção das imagens (sensor) estará sempre perpendicular à sua trajetória instantânea (vetor deslocamento);
2. A rua e as calçadas estarão, ao menos localmente, em um plano perpendicular ao sensor;
3. para todo plano paralelo ao sensor (e perpendicular à trajetória do mesmo) a escala será inversamente proporcional a profundidade, isto é, será constante em qualquer plano desta categoria⁶²;

⁶¹ Ao menos na maioria das imagens, há algumas exceções, como em esquinas e curvas.

⁶² Há uma demonstração gráfica, feita no aplicativo *geogebra*, disponível em: < <https://www.geogebra.org/m/neqv2wpb> >

Figura A2.1: uma típica imagem tomada com o SMMT



FONTE: O AUTOR (2019)

Assim sendo, devido a isso, basta conhecer alguma medida de antemão, para se obter o diâmetro de uma árvore (pode ser aproximada por um cilindro, cuja projeção em perspectiva será um retângulo, de modo que uma seção circular terá um segmento como imagem), por meio de uma simples relação de escala, de acordo com o demonstrado na equação A2.1.

$$\frac{DT}{DI} = \frac{MT}{MI} \therefore DT = \frac{MT}{MI} * DI \quad (A2.1)$$

Onde:

- DT: Diâmetro no Terreno (m)
- DI: Diâmetro na Imagem (pix)
- MT: Medida no Terreno (m)
- MI: Medida na Imagem (pix)

Vale ressaltar que tal medida e o diâmetro devem estar no mesmo plano vertical paralelo ao sensor, ou seja, também se assume que a árvore possui seu

tronco verticalizado. Tal técnica também só é propícia para o cálculo do diâmetro para árvores com o tronco uniforme.

Para a medida de referência, a medida da calçada é uma excelente opção, sendo que a idéia não é medi-la diretamente⁶³, mas sim obtê-la em locais onde sua medida já está disponível: mapas públicos; ortofotos; plantas; plano diretor, etc. Na figura A2.2 está apresentado um exemplo de como se realizou a medição das medidas em pixels da medida de referência (largura da calçada⁶⁴) e do diâmetro. As medidas no terreno foram feitas com o auxílio de uma trena.

Figura A2.2: Detalhe da realização das medidas de referência e do diâmetro na imagem



FONTE: O AUTOR (2019)

Na tabela A2.1 estão apresentados os resultados obtidos nos testes realizado com a técnica aqui apresentada.

⁶³ Vale lembrar que caso fosse necessário medi-la *in loco*, seria mais acurado medir diretamente a circunferência da árvore para se derivar o diâmetro.

⁶⁴ Na imagem foi medido um segmento pertencente a um plano paralelo ao sensor (linha horizontal na imagem, considerando que a feição imageada esteja em um plano perpendicular ao sensor), entretanto, no terreno foi medido o comprimento da calçada, que é um segmento perpendicular ao eixo da calçada. Ou seja: não se trata da imagem da medida no terreno a medida realizada na imagem. Esse é um erro sistemático, não corrigido no presente ensaio.

Tabela A2.1: Medidas indiretamente obtidas por meio de relações de escala

Árvore n	Medida Ref. (m)	Diâmetro Medido (m)	Medida Referência na Foto (pix)	Diâmetro na Foto (pix)	Diâmetro Calculado	Diferença (cm)	Diferença (%)
0	3.230	0.419	329	40	0.393	2.6	6.2
1	3.180	0.461	348	53	0.484	-2.3	-5.1
2	3.168	0.277	199	19	0.302	-2.6	-9.2
4	3.146	0.178	294	17	0.182	-0.4	-2.4
5	3.192	0.604	142	27	0.598	0.6	0.9
6	3.220	0.488	316	48	0.489	-0.1	-0.3
7	3.158	0.491	381	59	0.489	0.2	0.5
8	3.180	0.483	284	47	0.526	-4.3	-8.9

FONTE: O AUTOR (2019)

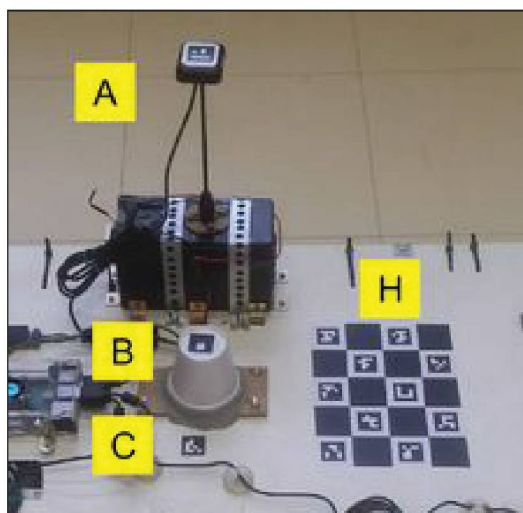
Pode se observar que mesmo diante da natureza aproximada da técnica, que envolve incertezas nas medidas, o resultado apresenta uma margem de erro em torno de 10%, sendo que em alguns casos há diferenças muito pequenas, de menos de 1%.

APÊNDICE 3 - ETAPAS NÃO IMPLEMENTADAS OU NÃO REPORTADAS

Como já comentado, o presente trabalho passou por uma reformulação. A demanda por uma série de procedimentos novos e prioritários, fez com que etapas em andamento ficassem de fora do trabalho. Contudo, isso não elimina sua relevância, nem mesmo o esforço realizado visando alcançá-las. Assim sendo, as mesmas serão aqui expostas.

1. Cálculo das *poses* relativas entre câmera/IMU/antena (iniciada, vários testes)
 - Como pode ser notado nas figuras em que a plataforma é retratada (sobretudo na 3.1), foram inseridos em locais estratégicos na plataforma uma série de alvos do tipo ARUCO (GARRIDO-JURADO, et. al., 2014). Tais alvos são de formato quadrado, como se conhece o lado dos alvos, é possível recuperar a escala da cena e determinar completamente suas coordenadas por meio de processos fotogramétricos. De posse das coordenadas tridimensionais dos alvos, por meio de geometria analítica é possível derivar as quantidades de interesse, como em VESTENA (2017). Na figura A3.1 está apresentado um detalhe contendo parte dos alvos presentes na plataforma, há juntamente aos alvos convencionais, um tabuleiro do tipo “*charuco*” (“*chessboard* + Aruco”), onde há pontos advindos das interseções de um tabuleiro, permitindo qualidade máxima devido à interpolação a nível subpixel, cuja finalidade será a de calibrar a câmera que for empregada para retirar as fotos da plataforma.

Figura A3.1: alguns dos alvos ARUCO da plataforma



FONTE: O AUTOR (2018)

2. Calibração dos parâmetros que descrevem a geometria da câmera da plataforma (implementada).
 - Etapa não empregada no *workflow* do trabalho como proposto atualmente, devido à pequena relevância no enfoque atual, ao contrário do tema anterior de pesquisa. As distorções geradas pelo sistema de lentes da câmera utilizada não possuem magnitude suficiente para afetar o desempenho da CNN empregada para classificação, a última calibração realizada, empregando apenas os parâmetros k_1 e k_2 do modelo de *Conrady-Brown*, devido à pequena magnitude das distorções, está disponível em um arquivo “.yaml” no repositório do *GitHub* apresentado no capítulo 3, no seguinte caminho: “*current_calib/camera/camerav2_1280x960_k1k2.yaml*”
3. Transformação espacial das poses para a fotografia (iniciada).
 - Uma vez obtidos os vetores e as matrizes de desalinhamento correspondentes ao item 1 dessa lista, será possível transformar a posição originalmente calculada para o CF da antena GNSS para o CP da câmara e a orientação calculada com relação ao BF da IMU para o BF da câmera;
4. Cômputo da solução integrada INS/GNSS para solução de navegação (iniciada)

- Há disponível para o ROS uma aplicação denominada “Robot Localization” (http://docs.ros.org/melodic/api/robot_localization/html/), contendo o filtro de Kalman. Assim sendo, uma vez dispondo das transformações do item 1, é perfeitamente possível implementar uma solução de navegação integrada;

5. Calibração da IMU (implementada e utilizada)

- Por meio de uma série de movimentos, especificados no manual fornecido pelo fabricante, é possível realizar a calibração dos sensores da IMU, e então armazenar os parâmetros em um arquivo “*.json*”, o principal parâmetro contemplado é o *offset* da IMU, isto é, a leitura média que o sensor registra ao invés de 0 quando em repouso. O Arquivo contendo a calibração empregada está disponível repositório do *Github*, apresentado no capítulo 3, com o seguinte caminho: “*current_calib/calib_imu.json*”.

6. Calibração Radiométrica para obtenção de valores de NDVI acurados (iniciada, vários testes realizados)

- Se faz necessário o emprego de alvos com valores de NDVI conhecidos. Idealmente devem ser alvos com resposta muito bem controlada (como um pedaço plástico produzido em larga escala industrial, ou lâmpadas/leds da cor em questão) em condições de iluminação muito bem conhecidas. Há uma larga fonte de material de como realizar tal procedimento em comunidades de entusiastas de imageamento multiespectral de baixo custo, nomeadamente o *PublicLab* (<https://publiclab.org/>). Há diversos artigos apontado para a mesma solução: deve-se alterar os ganhos e realizar o ajuste no balanço de branco separadamente das bandas 1 e 3. Entretanto isso deve ser feito antes da coleta dos dados, uma vez que tal configuração altera a forma como o sensor realiza as leituras, não havendo solução completamente satisfatória em pós-processamento. A não implementação de tal procedimento se ao fato que na versão atual do *driver* que integra à câmera ao ROS, não há suporte para ajuste manual do balanço de branco. Assim sendo, seria necessário que o

autor do presente trabalho implementasse por conta tal funcionalidade ou modificar toda a forma de captura de dados. Então o autor optou por priorizar a etapa de classificação empregando CNNs e por fim utilizou-se o balanço de branco no automático, de modo que os ganhos variam para compensar as mudanças de iluminação.

- Uma alternativa possível para obtenção de uma aproximação relativamente acurada seria o emprego de alvos com valores conhecidos na literatura geral: folhagem genérica saudável; folhagem genérica morta; asfalto... E então ajustar um modelo empírico, gerando uma curva de transferência da resposta resultante como função da resposta obtida;

7. Configuração da Antena GNSS (implementada e utilizada)

- Por meio de comunicação Serial, é possível configurar a antena GNSS, com uma série de opções, sendo necessário enviar tais comandos toda vez que o sistema é inicializado, já que a antena não possui memória não-volátil, dentre os comandos utilizados, se destacam:
 - Ligar o modo automotivo, informando ao filtro de posicionamento que a antena está em uma plataforma não-holonômica;
 - Ativar o uso das constelações GALILEO e GLONASS, para aumentar a quantidade de satélites utilizados;
 - Remover o envio de mensagens de navegação não utilizadas, que fariam o uso de banda desnecessariamente;

8. Materialização do Norte Geográfico nos dados de orientação;

- Atualmente, os dados de navegação geram uma orientação com eixo Y instantâneo apontando para o norte magnético local instantâneo. Após o processamento dos dados GNSS, pode-se interpolar uma declinação magnética (baseada em um modelo global) para cada instante e aplicá-la à orientação calculada pelo INS;

REFERÊNCIAS:

Garrido-Jurado, Sergio, et al. **Automatic generation and detection of highly reliable fiducial markers under occlusion**. *Pattern Recognition* 47.6 (2014): 2280-2292.

APÊNDICE 4 - SOBRE O TEMA ANTIGO E SEUS PROBLEMAS TÉCNICOS

A presente dissertação é produto de 30 meses de trabalho, dos quais a maior parte foi dedicada a um outro tema. O presente apêndice faz-se necessário, diante de possível questionamento a respeito do quão longe o autor foi capaz de chegar no trabalho apresentado por meio desta dissertação.

Anteriormente, o tema proposto era: “Desenvolvimento De Um Sistema De Navegação De Baixo Custo Baseado Em Odometria Visual-inercial E GNSS”, o qual foi aprovado no exame de qualificação, cujo texto apresentado à banca está disponível na íntegra no seguinte link: <https://rebrand.ly/qualificacao_antiga_kaue>. O texto pode ser utilizado como referência para compreensão do tema antigo, entretanto não elucida as razões pelo qual o mesmo foi abandonado em favor do presente, sendo este o assunto do presente Apêndice.

Vale lembrar que o presente trabalho não foi reiniciado “do zero”, a plataforma já estava fisicamente pronta⁶⁵, com os procedimentos de gravação de dados e realização de levantamentos, ambos já prontos, ainda que tenham demandado pequenas alterações. Contudo, toda a finalidade do trabalho passou por uma reproposição, acarretando na completa reescrita do texto, assim como se repensou os resultados a serem gerados, demandando uma reproposição em como gerá-los.

O tema antigo se tratava do emprego de todos os sensores presentes na plataforma para a obtenção de uma solução única de navegação de baixo custo. Para tanto, propunha-se o uso de um conjunto de sensores, basicamente os mesmos apresentados em 3.1, com a grande diferença de que no lugar da câmera aqui usada, eram usadas duas *Webcam Logitech C270*, cuja comunicação com o computador da plataforma se dava por meio de uma interface USB 2.0. Entretanto o computador usado, só possui uma única placa de suporte para tal protocolo, de modo que ambas precisavam dividir recursos. Com apenas uma de ambas as câmeras funcionando, já eram impostos atrasos nos dados que inviabilizam seu uso.

Após diversos testes com o comando *usbtop*, verificou-se que nenhuma câmera USB 2.0 seria capaz de atender aos requisitos da aplicação. Assim, substituiu-se a opção anterior de câmeras, pela atual, cuja integração com o sistema

⁶⁵ Foram apenas inseridos o filtro Vermelho e o anteparo *anti-glare*.

se dá por meio de uma interface especialmente projetada para o computador presente no sistema.

Ocorre que para uma aplicação envolvendo a odometria visual-inercial, é necessário que o tempo câmera e IMU obtenham (CHAO et. al. 2014) suas observações no mesmo exato instante de tempo⁶⁶. Há no sistema diversas fontes de erro que fazem com que tal condição seja violada, para exemplificar:

1. A câmera utilizada ainda que tenha uma comunicação especialmente projetada, ainda possui seu relógio próprio, diferente do relógio do computador, assim, seus sistemas de tempo são diferentes, não havendo algo voltado para sua sincronia. O mesmo ocorre para os demais equipamentos;
2. A câmera faz uma captura por linhas (*rolling shutter*), assim, há um instante de tomada para cada linha;
3. A IMU envia como saídas dados pré-processados e não os dados brutos. Acaba ocorrendo que tal pré-processamento é um processo que gera um pequeno atraso na transmissão dos dados;

Assim, devido a estes e outros erros com origem em questões de *hardware*, verificou-se a inviabilidade técnica da proposta. Para viabilização da solução do problema, seria necessária a realização de modificações físicas nos circuitos dos sensores, para se realizar uma *sincronização a nível de hardware*, possivelmente sendo necessário o uso de um único relógio para todos os dispositivos, ao menos para todos os sensores. Assim, conclui-se que seria necessária uma equipe multidisciplinar com ênfase em eletrônica, para tornar possível o êxito da proposta.

REFERÊNCIAS:

Guo, C. X., Kottas, D. G., DuToit, R., Ahmed, A., Li, R., & Roumeliotis, S. I. (2014, July). **Efficient Visual-Inertial Navigation using a Rolling-Shutter Camera with Inaccurate Timestamps**. In *Robotics: Science and Systems*.

⁶⁶ A tolerância se encontra numa ordem de grandeza muito pequena.

APÊNDICE 5 - MAPA PRODUZIDO COM OS DADOS PROCESSADOS

Um sistema de Mapeamento Móvel se propõe a gerar dados espaciais, que podem ser utilizados em bases digitais, assim como para a composição de mapas tradicionais.

Objetivando demonstrar tal capacidade, criou-se um mapa utilizando os dados coletados pelo SMMT. Vale se ressaltar que calcular a abrangência de maneira rigorosa, da cobertura vegetal baseada em tais dados exigiria processos elaborados, não contemplados na presente dissertação.

Assim, optou-se por uma abordagem simplificada: utilizar o percentual de pixels identificados como vegetação, em cada imagem como indicador da presença de cobertura vegetal. De fato, se trata de uma quantidade sem normalização ou padronização, havendo geralmente uma parcela relevante da imagem ocupada pelo asfalto e pelo céu. Não seria possível uma imagem com 100% de vegetação, ao passo que existem imagens com 0%. Então, o percentual de vegetação serve mais como indicativo do que como medida acurada do fenômeno analisado.

De posse dos percentuais relativos a cada foto, obtidos por meio do *forward pass* de cada imagem pela versão final da CNN empregada, a próxima etapa foi a obtenção da posição de cada foto. Por meio da interpolação Linear aplicada às posições determinadas pelo sistema de navegação, foi possível estabelecer a posição de cada fotografia tomada pelo sistema⁶⁷.

Assim, com o conjunto de pontos representados pelo conjunto de três informações: latitude; longitude; e o percentual em questão, criou-se a camada vetorial em formato “.csv”, para a composição do mapa no *software* QGIS. Utilizou-se como mapa-base a renderização padrão do *OpenstreetMap*, por ser uma fonte de dados completamente livre. Alterou-se a saturação de tal base, de modo a torná-lo mais acinzentado e não interferir na interpretação do tema de interesse.

Como trata-se de uma região muito pequena a ser representada, podem ser tomadas algumas liberdades na escolha da projeção cartográfica escolhida. Ainda assim, optou-se por criar uma projeção personalizada, de modo a que a

⁶⁷ Em uma abordagem simplificada onde se desconsidera o deslocamento entre o centro de fase geométrico antena GNSS e o centro de perspectiva da câmera.

representação da parcela do elipsóide considerada, seja feita com uma quantidade irrelevante de distorção. Então escolheu-se a projeção Transversa de Mercator, com Meridiano-Padrão -49° , de modo a posicionar o contato da superfície de projeção, bem próximo (cerca de 14' a leste) da área de interesse.

Dado que o retângulo envolvente dos dados de interesse possui uma maior extensão no sentido Norte-Sul, optou-se por um *layout* em orientação “retrato”. Utilizou-se (1:3600) a menor escala que permitisse a representação por completo da área de interesse, considerando apenas denominadores múltiplos de 100.

Por fim, para representar o tema apresentado, que se trata de uma variável quantitativa, utilizou-se a escala de cores *Magma*, por se tratar de um esquema de cores sequencial com uniformidade perceptual.

O mapa produzido encontra-se apresentado na próxima página, estando também disponível para *download* no repositório⁶⁸ do projeto da presente dissertação, no diretório “*map*”, onde estão disponibilizados os demais arquivos para reprodução de tal mapa, como o projeto para QGIS e o arquivo de *layout*.

REFERÊNCIAS:

OpenStreetMap contributors. **OpenStreetMap**. Disponível em: < <https://www.openstreetmap.org> >

SISNEROS, Robert et al. **Chasing Rainbows: A Color-Theoretic Framework for Improving and Preserving Bad Colormaps**. In: International Symposium on Visual Computing. Springer, Cham, 2016. p. 391-402.

⁶⁸ disponível em: < <https://github.com/kauevestena/snay> >

PERCENTUAL DE PIXELS DE VEGETAÇÃO NAS IMAGENS TOMADAS PELO SMMT DO LAPE (UFPR) NO CENTRO POLITÉCNICO DA UFPR

